External Cavity Quantum Cascade Lasers

D i s s e r t a t i o n

zur Erlangung des akademischen Grades

d o c t o r r e r u m n a t u r a l i u m

(Dr. rer. nat.)

im Fach Physik

eingereicht an der

Mathematisch-Naturwissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

Dipl. Phys., Jan-Ferenc Kischkat

Präsidentin/Präsident der Humboldt-Universität zu Berlin

Prof. Dr. Jan-Hendrik Olbertz

Dekanin/Dekan der Mathematisch-Naturwissenschaftlichen Fakultät

Prof. Dr. Elmar Kulke

Gutachter/innen: 1. Professor W. Ted Masselink

2. Professor Heinz-Wilhelm Hübers

3. Professor Andrew Holmes

Tag der mündlichen Prüfung: 31. 08. 2015

This work is dedicated to my mother, my father, Till, and Trevor.

0.1 Acknowledgements

I want to thank Professor Masselink and Dr. Semtsiv for being the most helpful and inspiring advisors one could ask for. Whenever physics-related help and advice was needed, it was on point, whenever general questions of direction arose, they were taken on with creativity and dedication, and whenever a friendly chat was all it took for motivation, the time of day was spent freely. Thank you very much for this.

I want to thank my committee members Professor Hübers from the DLR, Professor Holmes from Imperial College, Professor Riechert from the Paul-Drude-Institute, and Professor Nolting from Humboldt University for reviewing this ample piece of work thoroughly.

I also want to thank Sentech Instruments, particularly Mr. Peters, for repeated help and support with the ellipsometric measurements and greatly inspiring expertise.

My dearest thanks goes to my mother Anna Tóth-Kischkat, my brother Till-Árpád Kischkat, and my stepfather Trevor Semple, for all of their love, their general support, and for their rediculously effective help on a myriad of particular occasions. Trust me, I do not take it for granted. Even when I am not in the state of mind to show the appropriate appreciation in the moment. I love you guys.

And I want to thank the wonderful Miss Priyanka Sharma for being around throughout the entire time it took to do this. I believe she had more work with this thesis than I did, but her patience, support, and encouragement somehow never ended. Thank you so much for this!

0.2 Abstract

This thesis thoroughly investigates theoretically and experimentally the effects many physical parameters have on the performance of Tunable External-Cavity Quantum-Cascade Lasers (EC-QCLs). These include, among others, the anti-reflection coating, the type of optics, and the geometrical as well as mechanical and structural properties of the EC setup. This was done by assembling three very different EC setups and comparing and discussing their performance, as well as advantages and disadvantages for different purposes using mainly QCLs from the same original wafer for better comparability.

For the last part of this thesis, a new type of EC-QCL configuration was developed with properties so promising that we believe it has the potential to replace the Littrow Cavity in the long term. This is an alignment-stabilized and interference filter-tuned design using a retroreflector as the external reflector. For the demonstration of this concept, development of the tuning element in the form of an angle-tunable high-Q mid-infrared bandpass filter was necessary. For the design of the filter, materials with very strict tolerances on the physical and optical properties were selected from theoretical considerations and a fabrication method with highly optimized process parameters was developed. The first filters on the basis of yttrium fluoride/yttrium oxide/germanium/silicon have a transmission bandwidth of 0.14% of the central wavelength and a peak transmission of approximately 60%. The EC configuration resulted in a sensitivity reduction to mechanical perturbations of the reflector by two orders of magnitude, with a calculated potential for three orders of magnitude using optimized optics. This design lifts the fundamental constraint on miniaturization imposed on the Littrow design that requires large beam diameters to ensure a small bandwidth of the Littrow grating.

Keywords: EC-QCL, Anti-reflection Coating, Infrared Spectroscopy, Interference filter

0.3 Zusammenfassung

Diese Arbeit untersucht den Einfluss verschiedener physikalischer Parameter auf das Verhalten von Frequenz-abstimmbaren External-Cavity Quantenkaskadenlasern (EC-QCLs) theoretisch und experimentell. Diese beinhalten unter anderen die Antireflexschicht, die Art der Optiken, die geometrischen und die mechanisch/strukturellen Eigenschaften. Dies wurde erreicht durch Aufbau dreier sehr unterschiedlicher EC-Konfigurationen, der Diskussion und dem Vergleich ihrer Leistungsmerkmale und ihrer Vor- und Nachteile für verschiedene Anwendungen unter hauptsächlicher Verwendung von QCLs desselben Wafers der Vergleichbarkeit wegen.

Für den letzten Teil dieser Arbeit wurde ein neuer Typus EC-QCL mit vielversprechenden Eigenschaften entwickelt, sodass wir glauben er hat das Potential das Littrow Design langfristig abzulösen. Dieses selbststabilisierende Design verwendet einen Retroreflektor als externen Reflektor. Für die Demonstration dieses Konzepts war die Entwicklung eines Tuning-Elements in Form eines Winkel-verstimmbaren Mittinfrarot-Bandpass-Interferenzfilters mit

sehr hohem Gütefaktor vonnöten. Für das Design des Filters wurden Materialien mit sehr strengen Toleranzen bezüglich ihrer physikalischen und optischen Eigenschaften auf Basis von theoretischen Überlegungen ausgewählt und eine Fabrikationsmethode mit hochoptimierten Prozessparametern entwickelt. Die ersten Filter auf Basis von Yttriumfluorid/Yttriumoxid/Germanium/Silizium haben eine Transmissionsbandbreite von 0.14% der Zentralwellenlänge und eine maximale Transmission von etwa 60%. Die EC Konfiguration resultierte in verminderter Empfindlichkeit gegenüber Mechanischen Störungen des Reflektors um zwei Größenordnungen. Das Design behebt die grundsätzliche Limitierung des Littrow Designs bezüglich Miniaturisierung, da kein großer Strahldurchmesser vonnöten ist um kleine Bandbreiten des Littrowgitters zu erreichen.

Keywords: EC-QCL, Antireflex-Beschichtung, Infrarot-Spektroskopie, Interferenzfilter

Contents

0.4 List of Symbols

Latin Symbols:

a_i	weight factors of different laser transitions
A	facet area
A_{21}	is the Einstein A Coefficient of the transition
b	width of the QCL waveguide
B_{21}	is the Einstein B Coefficient of the transition
c	speed of light in vacuum
d	(Chapter 1) the groove spacing of diffraction grating
d	(Chapter 2) thickness of 1-layer AR coating
d_i	thickness of layer i in layer stack
d_H	thickness of high-index layer
d_I	thickness of initial layer
d_L	thickness of low-index layer
$D(\nu_0)$	dispersion
$E(z,t)$	position and time-dependent (complex) magnitude of the electric field vector
E_0	value of the electric field vector at $t = 0$ and $z = 0$
f	focal distance of lens
f_m	optical frequency of light of mode ν_m
F	Finesse of FP etalon
$F(x)$	normalized electric field profile, coordinate x is scaled with the beam half-width on the grating (the full beam thus extends from $x = -1$ to 1 with a width of 2)
$\bar{F}$	feedback from external cavity
g	gain coefficient as a function of wavenumber
$g(\nu)$	gain coefficient
$\bar{g}$	gain coefficient divided by pump current, constant
h	Planck's constant
$\hbar$	reduced Planck's constant
I	pump current
I_i^0	pump current with zero detuning of feedback element i
I_i''	second derivative of pump current with zero detuning of feedback element i
$\hat{I}(\nu)$	light intensity
J	pump current density
J_{th}	threshold current density
k	extinction coefficient
k	(Chapter 1) lattice vector
$k_{\|\|}$	in-plane component of the lattice vector (normal to growth axis)
$\hat{k}$	complex magnitude of electric field vector

l	length of the QCL waveguide
$l(\nu)$	is the normalized lineshape function of the transition
$\hat{l}$	(Chapter 5) effective distance light travels in FP cavity
L	effective free-space path length between the intra-cavity facet and grating
L_p	thickness of one period of active region
M_{21}	transition Matrix element
n	is the index of refraction
n_{eff}	effective modal index of refraction of the waveguide
$n_{\text{eff}}(\nu_0)$	is the effective index of refraction of the waveguide at the of the mode with central wavenumber ν_0
$n_{\text{g}}(\nu_0)$	group index of refraction
n_s	substrate index of refration
n_a	ambient index of refration
n_H	refractive index of high-index material
n_L	refractive index of low-index material
N	population difference
N_1	density of electrons in the lower state
N_1	density of electrons in the upper state
N_0	is the steady-state solution to N in the absence of radiation
$\tilde{N}$	number of round trips of light in FP cavity
$\hat{N}$	number of round trips of light in FP cavity for power to decay to $1/e$
$\bar{N}$	number of layer pairs in DBR
P	output power
P_i	output power of facet i
$P(t)$	power within FP cavity as a function of time
$\tilde{P}$	E-field polarization (TE or TM)
q	symbol for $2d\sin(\theta)/\lambda$
Q	quality factor of FP etalon
$r_1^{(+)}$	Fresnel amplitude reflection coefficient of the extra-cavity facet, forward direction
$r_2^{(\pm)}$	Fresnel reflection coefficients of the intra-cavity facet, forw. + backw. directions
r_i	amplitude reflection coefficient reflectors i
$\hat{r}(\nu, \nu_g)$	complex effective amplitude reflection coefficient of the External Cavity
$\vec{r}$	position vector
R_0	reflectivity of uncoated facet
R_i	reflectivity of facet i
r_G	grating reflection coefficient
R_g	grating reflectivity
$\hat{R}_i$	pump rates into level i
$\hat{R}$	pump rate
s	slope efficiency
s^j	slope efficiency of light from facet j

$\bar{s}$	width of the grating diffraction efficiency near the blaze angle θ_B
t	time
$t_2^{(\pm)}$	Fresnel transmission coefficients of the intra-cavity facet, forw. + backw. directions
t_{sp}	spontaneous emission lifetime of the transition
T	temperature
T_0	characteristic temperature of the QCL relating to threshold
T_{rt}	round-trip time of light in FP cavity
$U(\lambda, \theta_B, \tilde{P})$	effective diffraction function of each grating tooth
W	probability of absorption or stimulated emission of a photon by an electron
$\tilde{W}(m)$	normalized amplitude of the electric field at the mth grating tooth
x	(Chapter 5) fractional power loss of light during round trip in FP cavity
X	symbol in analytical result of stack reflectivity
$\bar{X}$	ratio of laser threshold current density before and after coating
$\bar{Y}$	ratio of laser slope efficiency before and after coating
z	coordinate coaxial with the waveguide
z_0	effective thickness of first-order FP cavity
z_{32}	optical dipole matrix element connecting the upper and lower laser states

Greek Symbols:

α	sum of the laser losses
$\bar{\alpha}$	effective loss (including overlap factor), equal to threshold gain during laser operation
$\alpha(\nu)$	sum of the laser losses as a function of wavenumber
α_w	distributed waveguide loss due to scattering
α^{EC}	External Cavity resonator loss
α^{FP}	Fabry-Perot resonator loss of QCL chip
α_{mi}	mirror loss of mirror i
$\beta(\nu)$	propagation constant
γ	gain of the laser amplifier
$\gamma_0(\nu)$	unsaturated gain, or small-signal gain
γ_{32}	linewidth of the laser transition (in units of energy)
Γ	confinement factor
δ	detuning angle of filter or grating
δ_k	phase symbol of layer k in AR coating
ΔE_C	conduction band offset
$\Delta\nu$	width of the lineshape function of the transition
$\Delta\nu_g$	Tuning Range of the External Cavity
ε_0	vacuum dielectric constant
η	net (amplitude) coupling efficiency each way due to the intra-cavity optics
η_d	external differential quantum efficiency
η_G	diffraction efficiency at the blaze angle

η_i	internal quantum efficiency
η_e^j	extraction efficiency of facet j
θ	External Cavity tuning angle (Littrow or filter)
θ_B	grating blaze angle
θ_{design}	design tuning angle of interference filter
λ_0	wavelength in vacuum
λ	wavelength
λ_{design}	design wavelength of interference filter
ν_m	wavenumber of laser or FP mode with index m
ν_0	position of maximum of gain curve
$\nu_g(\theta)$	Littrow wavenumber at tuning angle θ
ξ_{sp}	phenomenological term for spontaneous emission
$\hat{\xi}_{sp}$	spontaneous emission coefficient divided by pump current, constant
$\Xi(\kappa)$	spatial Fourier transform of the beam profile on the grating
$\sigma(\nu)$	transition cross section of absorption and spontaneous emission
τ_i	overall lifetimes of carriers in levels i
τ_{ij}	is the inverse of the scattering rate of electrons from level i to j
τ_s	effective lifetime symbol
ϕ	internal photon flux density
$\phi(\nu)$	internal photon flux density as a function of wavenumber
$\phi_1(z)$	density of photons traveling in the positive z direction
$\phi_2(z)$	density of photons traveling in the negative z direction
ϕ_i	phase symbol of layer i
$\phi_s(\nu)$	saturation photon flux density of the active region
$\hat{\phi}$	phase factor of 1-layer AR coating
$\hat{\phi}_i$	phase factor of layer i in AR coating
ω	frequency of photon
ω_0	width of the beam at the mirror of the cat's eye
ω_g	width of the beam at the grating
ω_{ic}	width of the beam within the External Cavity
Ω	frequency of phonon

Special Symbols of Matrix Method, Chapter 2.2:

d_k	thickness of medium k ($k = 1...p-1$)
E_p^+	amplitude of electric field vector in incident beam
E_p^-	amplitude of electric field vector in reflected beam
E_0^+	amplitude of electric field vector in transmitted beam
E_k^+	amplitude of electric field vector in beam traveling in positive direction in medium k ($k = 1...p-1$)
E_k^-	amplitude of electric field vector in beam traveling in

	negative direction in medium k $(k = 1...p-1)$
M_k	transmission matrix of layer k
n_k	refractive index of medium k $(k = 0...p)$
p	layer counter Matrix Method
r_p	reflection coefficient of layer stack, parallel component
r_{stack}	reflection coefficient of layer stack
R	Reflectivity of layer stack
R_p	Reflectivity of layer stack, parallel component
R'_p	Reflectivity of layer stack, perpendicular component
T	Transmissivity of layer stack
T_p	Transmissivity of layer stack, parallel component
T'_p	Transmissivity of layer stack, perpendicular component
v_k	Matrix element symbol
x_k^+	Matrix element symbol
x_k^-	Matrix element symbol
x_0^+	Matrix element symbol
x_p^+	Matrix element symbol
x_p^-	Matrix element symbol
u_k	Matrix element symbol
y_k^+	Matrix element symbol
y_k^-	Matrix element symbol
y_0^+	Matrix element symbol
Y_k^+	y component of E_k^+ $(k = 0...p)$
Y_k^-	y component of E_k^- $(k = 0...p)$
Z_k^+	z component of E_k^+ $(k = 0...p)$
Z_k^-	z component of E_k^- $(k = 0...p)$
β_k^+	y component of magnetic field vector corresponding to E_k^+
β_k^-	y component of magnetic field vector corresponding to E_k^-
γ_k^+	z component of magnetic field vector corresponding to E_k^+
γ_k^-	z component of magnetic field vector corresponding to E_k^-
κ	is a normalized vector denoting the direction of propagation
ϕ_k	angle of incidence in the positive direction in medium k $(k = 1...p-1)$
ω_k	Matrix element symbol

Special Symbols of Heat Exchanger Design, Chapter 4.10:

a	height of the channel
a_1	length of the area to be cooled
a_2	width of the area to be cooled
A	cross sectional area of the channel
A_c	cooling area

b	width of the channel
c_p	isobaric heat capacity
D_H	hydraulic diameter of channel
f	Darcy-Weisbach friction factor
h	heat transfer coeffcient
j	mass flux
k	thermal conductivity of the coolant
k_b	bend coefficient
l	length of the duct
m	number of bends
n	number of parallel channels
$\tilde{n}$	symbol
Q	volumetric flow rate
R_b	bend radius
R_e	Reynolds number
u	flow velocity
u_*	shear velocity of the coolant
v	speed of the coolant
W	overall heat exchange
x	filling factor of cooled area
y	coordinate normal to the wall
δ_ν	laminar sublayer thickness
ΔP	pressure drop across channel
ΔP_1	pressure drop in straight tube
ΔP_2	pressure drop from bend
ϵ	length scale of pipe roughness
θ	bend angle
ν	kinematic viscosity of coolant
ρ	density of the coolant
τ	shear stress within the water flow and

1

Fundamentals

1.1 Introduction and Motivation

Infrared spectroscopy is the investigative method with the greatest potential to deliver a compact and affordable trace gas sensor of very high sensitivity and specificity to a multitude of molecules simultaneously.

The mid-infrared (MIR) part of the electromagnetic spectrum spans, depending on definition, the wavelength range of approximately 3-12 μm, which corresponds to photonic energies of 100-400 meV. This range contains the excitation energies of the fundamental modes of vibration for a plethora of large and small, organic and inorganic molecules [2].

Since for most molecules the vibrational modes are *infrared active*, i.e. these molecules have either permanent or induced electric dipoles, this results in strong absorption of light passing through a set of molecules if its energy equals the energy of a vibrational excitation. For free molecules, fine-structure is added to the main vibrational absorption line due to a simultaneous transition in rotational state during the vibrational excitation. This is called a *rotational-vibrational* or *ro-vibrational* excitation. Infrared spectroscopy can thus be used to study the ro-vibrational structure of free molecules. This structure is unique to most molecules. As a consequence, if the structure has been recorded once, the absorption spectrum can be used to positively identify the presence of a molecule. This is why parts of the MIR are often termed the "fingerprint region" of these molecules.

Due to the amount of information present in a broadband absorption spectrum, it is even possible to analyze mixtures of analytes and accurately determine the identity and quantity of the components. This is simply done by numerically decomposing the complicated spectrum of the mixture into the spectra of the individual types of molecules. Due to the strong interaction with the light, the method has very high sensitivity to even the smallest quantities of the molecules, which can be further enhanced by making use of appropriate multi-pass sample cells that utilize mirrors to reflect the light many times through the analyte, thereby multiplying the probability for absorption.

The high sensitivity and specificity make infrared spectroscopy interesting for detection and quantification of trace gases in an unknown and uncontrolled mixture such as ambient

air or human breath. It is known that many diseases such as lung cancer or diabetes produce substances that are naturally exhaled through breath in small quantities [3]. Detecting these substances can be a very useful tool to non-invasively diagnose these diseases. Ideally, detection is sensitive enough to diagnose the disease in an early stage, where it can easily be cured with conventional methods.

Infrared absorption spectra are not only unique to chemically identical molecules, but also to the isotopic composition of the molecules, i.e. different *isotopologues*. Thus, it is even possible to determine the isotopologue composition of a set of nominally identical molecules. This effect can be used for many interesting diagnostic techniques. For instance the human metabolism can be probed by incorporating a dose of isotopically marked atoms or molecules, such as carbon-13, and then tracing these by measuring the elevation of isotopically marked CO_2 in the exhaled breath over time.

Infrared spectroscopy has many practical configurations, that mainly split up into *dispersive* techniques and *Fourier transform* techniques. In dispersive techniques, a wavelength is scanned through the range of interest and traverses the sample one wavelength at a time. For all light sources other than lasers, this means filtering out only the desired wavelength (with a narrow band around it) from the emission with the use of a monochromator and tuning it, discarding the remaining power outside this range (these will be called "passive" dispersive techniques here).

In Fourier transform infrared spectroscopy (FTIR) [4], however, all wavelengths pass through the sample at the same time. Here light from a broadband source is split into two beams, one of which travels a periodically varying distance through the spectrometer as it is reflected by a moving mirror. After recombining the two beams, they are passed through the sample. The intensity as a function of mirror position is then recorded as an interferogram, which then also contains the absorption information of the sample. Upon Fourier transforming the interferogram, the absorption spectrum of the sample is revealed.

FTIR spectroscopy has been the dominating technique of infrared spectroscopy for the longest of time. This is due to the fact that for decades the only practically available MIR light source for spectroscopic applications was the Globar (TM), which is a heated bar of silicon carbide emitting blackbody radiation in the infrared. As a thermal broadband source running at a moderate temperature, the Globar has quite a low level of light emission and an exceedingly small spectral radiance.

But FTIR spectroscopy has two great advantages inherent to its design that greatly reduce the problems associated with the low (spectral) radiance of the Globar: *Fellgett's advantage* states that the achievable signal-to-noise ratio (SNR) is higher for a multiplex technique such as FTIR (where all wavelengths pass through the sample simultaneously) than for passive dispersive techniques using the same source. *Jacquinot's advantage* states that the available light throughput is higher than for passive dispersive techniques using the same source, since it is only determined by the diameter of the collimated beam within the FTIR spectrometer. Whereas in passive dispersive techniques, the entrance and exit slits of the monochromator result in a tradeoff between available light levels and linewidth.

If the Globar is the only light source available, these advantages turn out to be crucial in achieving any acceptable level of sensitivity to lower concentrations of analytes – especially if they have narrow absorption lines. This is ultimately the reason why the development of FTIR spectrometers is so far advanced that it almost seems impossible for other techniques to "catch up".

FTIRs, however, have some great disadvantages, namely their great complexity, which results in considerable size and cost, as well as their sensitivity to any type of environmental perturbation. They are also not easily integrated with multi-pass gas cells. For all these reasons, FTIR spectrometers are not very promising in widespread use in any environment other than a specialized laboratory.

Tunable laser sources work fundamentally different to thermal sources. They concentrate *all* of their available power in a single emission line that, depending on the exact laser, is 5-8 orders of magnitude narrower than the bandwidth of, for instance, the Globar. However, they can have similar overall power levels resulting in a hugely increased spectral radiance. Their output is also normally concentrated in a narrow collimated beam that has a 1-3 orders of magnitude smaller cross-section than the beam within the FTIR, which further increases the radiance. So although tunable laser sources are inherently dispersive in nature, their advantages in radiance quite obviously outweigh the SNR-increasing advantages of the FTIRs with a Globar (by many orders of magnitude). In addition, since dispersive techniques are normally much simpler in build and are better suited for the integration with multi-pass gas cells and many other detection methods, these "active" dispersive techniques using tunable lasers have substantial practical advantages for any application of infrared spectroscopy, making the pursuit of these very promising, despite the FTIR's phenomenal head start.

For the longest of time, tunable laser sources in the MIR were few and far between, because the MIR is generally a range that is not easily accessible with lasers [5]. On the one hand, only very few (doped insulator) solid-state lasers have emission lines in the MIR, but these are not widely tunable [5]. Semiconductor interband lasers emitting in the MIR, on the other hand, must necessarily be made of materials with a very small band gap. These are typically ternary or quarternary lead salt compounds. But these materials have very poor mechanical properties that severely limit lifetime of these lasers [5]. Also, they have very low efficiencies and need to be run at cryogenic temperatures.

Due to this limited availability and usefulness of lasers with direct emission in the MIR, the most practical sources were ones that convert the emission of a laser of a different wavelength (e.g. in the near-infrared) into the MIR using non-linear crystals. This is done either by difference-frequency generation of two laser sources [5] or using a single source in an optical parametric oscillator (OPO) [5]. However, since these use multi-photon processes that rely on very high photon densities within the non-linear crystal, they can only be run efficiently at very short pulse durations.

Lately, however, MIR spectroscopy is on its way to a revolution with the increasing availability of the *Quantum Cascade Laser* (QCL) (photographed in Fig. 1.1). This compact and rugged semiconductor laser combines all the advantages of the diode laser (DL) in terms of

footprint and reliability – which has revolutionized telecommunication and specialized illumination in the visible (VIS) and Near-Infrared (NIR) ranges – with direct emission in the Mid-Infrared (MIR). Contrary to interband diode lasers, laser action in QCLs is not based on interband recombination of electrons and holes, but on *unipolar intersubband transitions* in a structure of alternating quantum wells and barriers. For this reason, well-controllable and reliable wide-bandgap materials can be used that do not exhibit the typical drawbacks of small-bandgap materials [5]. Through variation of the quantum well thicknesses, and thereby the energy levels of the confined electronic states between which the laser transition occurs, the emission wavelength can be freely placed in an exceptionally wide range that contains the entire fingerprint range between 3 and 12 μm and reaches as far as the THz regime.

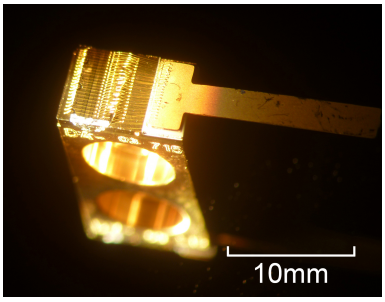


Figure 1.1 *Photograph of a QCL on a gold-plated copper heatsink.*

Since the electrons do not leave the conduction band during such transitions, the process can be cascaded and the electrons essentially "recycled", resulting in high quantum efficiencies (>1) and finally, high powers. Since the wavelength is determined by the design of the structure and not the materials, blocks of cascades of different designs can be grown on top of each other in the same material system [6]. By stacking structures with different central wavelengths and overlapping gain, active regions with very broad gain can be produced. Alternatively, structures can be designed to have multiple upper laser levels and/or multiple lower laser levels, which also has the effect of broadening the gain [7]. The extreme case of the latter are superlattice QCLs, in which both the upper and lower laser levels are formed of "minibands", which are the (quasi-continuous) Bloch states of the superlattice in analogy to the valence and conduction bands of the unstructured crystal.

The most basic embodiment of the QCL is the Fabry-Perot (FP) laser, where the optical feedback necessary for resonant stimulated emission is caused by Fresnel reflections from the cleaved ends, or facets, of the active region waveguide. Since this reflection is non-selective to the wavelength, FP lasers are inherently *multimode* and especially when driven with short current pulses, they emit over a considerable part of the gain region simultaneously. For dispersive spectroscopy, however, it is necessary to narrow the emission to a line fine enough to resolve the finest structure of the absorption under investigation, best to a single laser mode. This mode then needs to be tuned over the desired range.

There are two main strategies to achieve this. Tuning the emission over a narrow range, typically a few cm^{-1}, can be achieved monolithically using distributed feedback (DFB) QCLs [8]. Here, optical feedback is caused by a Bragg grating etched into the top of the waveguide.

This forces the emission wavelength to coincide with the effective period of the grating. The wavelength is then tuned by varying the effective grating pitch through the chips's refractive index. This in turn is varied through the temperature of the medium either by directly controlling the temperature of the heat sink or by driving the laser with short current pulses. The rapid periodic heating of the active region during the pulse (with subsequent cooling between pulses) then chirps the wavelength with great speed [8].

To truly compete with FTIR, however, the ability to tune the wavelength over a much wider spectral region is necessary. This can be done using a so-called *External Cavity* (EC) configuration, where feedback is caused by an external reflector with a strongly wavelength-selective reflectivity. This forces laser oscillation into the narrow band selected by the reflector. The most common configurations are the so-called *Littrow* and *Littman-Metcalf* configurations that both employ a diffraction grating as the wavelength-selective element. By tuning the angle of incidence of the beam impinging on the grating, e.g. by tilting the grating, the emission wavelength of the laser can thus be tuned.

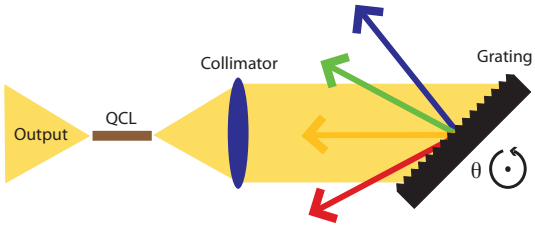


Figure 1.2 *Sketch of a Littrow-type external cavity.*

A sketch of the most commonly used *Littrow External Cavity* configuration is given in Fig. 1.9. Since the output beam of a semiconductor laser amplifier such as a QCL is highly divergent, it has to be collimated, as is done with a lens in this sketch. The collimated beam then impinges on the diffraction grating, which disperses the light according to wavelength. If the grating is correctly aligned, one narrow frequency band is aimed straight back at the lens, on the same path as the incoming beam, and is focussed into the ("intra-cavity") laser facet. This causes a resonant feedback loop for the light between the grating and the QCL's other ("extra-cavity") facet with the QCL acting as the light amplifier.

Due to the relative ease of alignment of the Littrow setup in its simplest form, resulting from the small number of components – a small lens and a diffraction grating – as well as the off-the-shelf availability of these, it has become the dominant External Cavity configuration, holding almost a monopoly within all spectroscopic applications that employ EC-QCLs. For the same reason, the first company developing commercial devices was founded only four years after the first pioneering publications of QCLs in Littrow ECs. These companies, mainly spear-headed by Daylight Solutions in Poway, CA, USA, and Pranalytica Inc. in Santa Monica, CA, USA, now supply the academic world with increasingly complex turn-key solutions for their spectroscopic applications, however staying true to their core technology. This results in only very few academic efforts to rethink and optimize the basic design concepts due to the great head start of these companies.

This is, in part, unfortunate, since the Littrow configuration inherently has some severe drawbacks, most of all its mechanical fragility due to the extreme alignment accuracy of the components, required for the resonant feedback to be maintained. Thus the Littrow design is not a design made for small portable applications. In order to achieve some level of stability, the commercial devices employ ultra-stiff structures and ultra-stable yet highly accurate positioning systems, that consequently make these setups rather costly and bulky, while still not being overly useful for applications outside a protected laboratory environment.

Finally, while research on the QCLs on the one hand, and research on a plethora of spectroscopic methods and applications on the other, is ongoing and vivid, very little has been published on the basic interactions within the External Cavities. Thus, aside from a few very interesting ideas, not many efforts have been made in the pursuit and investigation of different EC concepts that could potentially resolve some of the fundamental drawbacks of the Littrow configuration. Here we saw potential for new research, i.e. in the better understanding of the Littrow concept itself in order to optimize geometric and other parameters – i.e. parameters not intrinsic to the QCLs such as for instance the shape of the gain profile – as well as to put forth new ideas for completely different designs of ECs.

This thesis thoroughly investigates theoretically and experimentally the effects many physical parameters have on the performance of EC-QCLs. These include, among others, the anti-reflection coating, the type of optics, and the geometrical as well as mechanical and structural properties of the EC setup. We have done this by assembling a number of very different EC setups and comparing and discussing their performance, as well as advantages and disadvantages for different purposes using mainly QCLs from the same original wafer for better comparability.

For this thesis, various different External Cavity configurations have been developed with very different properties that make them useful for different applications. Their underlying interactions and properties are extensively investigated experimentally and theoretically, where possible using the same QCL for comparability.

The remainder of this Chapter 1.1 treats the fundamental theory of EC-QCLs.

Chapter 2 thoroughly treats the antireflection (AR) coating of the intra-cavity Laser facet used to reduce the detrimental effect of partial reflectivity of this facet on the overall performance of the External Cavity. This topic is not documented well in the literature. Normally, the coated QCLs are treated like lasers with a severely reduced reflectivity of one of the mirrors. However, when one of the mirrors of a laser is essentially removed, the laser seizes existence as a laser and behaves more like a light-emitting diode (LED), emitting only spontaneous emission, or a super-luminescent diode (SLD), emitting mainly amplified spontaneous emission. We have developed a model to better describe the transition region of LED, SLD, and laser for a laser with a good AR coating. This model uses elements of fiber amplifier theory and proved very useful for determining many parameters of the laser, including the actual residual reflectivity of the AR coating, from the output power–current characteristic of the laser before and after coating. We have also developed a method to deposit high-quality AR coatings using magnetron sputter deposition with an in-situ monitoring process of the laser

emission, the most direct way to determine the quality of the coating during the deposition process.

Chapter 3 demonstrates a very versatile Littrow External Cavity setup using only reflective optics and a visible pilot beam for alignment. This setup is a useful research tool, since it allows the use of nearly any experimental semiconductor laser even with very low emitted power or the necessity for cryogenic cooling, emitting at any wavelength from the Terahertz regime to the Ultra-Violet. This setup uses a visible pilot beam for alignment and can therefore be used to align QCLs inside an EC without using the beam of the QCL itself. This is very useful if the laser – particularly after depositing an AR coating – emits very low power levels or needs cryogenic cooling. This setup has lead to patent no. DE102012000038A1 [1].

Chapter 4 demonstrates a mechanically stabilized Littrow setup, which due to the great reproducibility of its experimental results can be used to investigate and compare the performance of the employed QCL in a number of different conditions. A turn-key tunable laser source, with various different I/O interfaces, complete with driving software and thermo-electric cooling has also been developed around this setup to perform complex spectroscopic experiments, some preliminary results of which are also included. A conventionally water-cooled heat sink to extract the heat from the high-power thermo-electric cooler has also been developed from first principles and shows performance better than any heat sink available on the market. This laser has already successfully been used by other research groups to perform absorption measurements, e.g. pump-probe measurements using an ultra-violet laser to excite the dye Coumarin 314.

Chapter 5 presents the first experimental demonstration of a fundamentally new External Cavity design that eliminates the problem of alignment sensitivity. It does this by being in essence "self-aligning", since EC feedback is realized by a retroreflector that has the property of reflecting the beam back to where it came from, thus automatically focussing the light back into the laser facet. Due to its self-stabilizing nature, the setup can not be misaligned without physically damaging the structure. Wavelength selection is achieved by an ultra-narrowband tunable interference filter. Since filters that meet the requirements for laser line selection have not existed prior to this work, we have developed one. For this we have thoroughly investigated potential materials and deposition techniques to be used for this cause and developed a process with strictly optimized and monitored parameters that produces filters with a finesse of ≈ 700. The filter thus has a similar wavelength selectivity as a Littrow grating and could therefore be used to demonstrate a prototype of this External Cavity setup. Aside from great mechanical stability, this EC configuration has many other advantages discussed in detail in that chapter. It can be miniaturized to a far higher extent than is allowed for the Littrow setup for physical reasons. It also allows for the use of positioning mechanics that are far more cost-effective than in current Littrow setups without performance degradation. And finally there is ample potential for cost-reduction through the mass-producibility of the components in use. For these reasons we believe this setup has the potential to one day fully replace the Littrow configuration.

Appendix A is a thorough reference-style overview of all the interactions contributing to the

dielectric properties of solids. The chapter explores these interactions in order to investigate from first principles which materials are "useful" materials for infrared optical coatings. High-quality optical coatings is a main focus of this thesis, used for anti-reflection coatings and bandpass interference filters, and very few material lend themselves to this use in the mid-infrared. Since infrared optical coatings are far less advanced than coatings in the VIS-NIR and quite little is published on the topic specifically, we have chosen to include a thorough overview into this thesis to serve as a practical manual for researchers getting involved in the field. During the development of this chapter we have found what we believe to be an entirely new, simplified but accurate theoretical model for the refractivity of solids, underlined with a thorough survey of common textbook misconceptions on the topic.

Appendix B provides the theoretical and experimental details of Spectroscopic ellipsometry, a method of investigating optical properties of thin films and characterizing thin film structures and used plenty of times throughout this thesis.

Appendix C gives the numerical programs used to fit laser parameters, including the residual reflectivity of AR coating according to the model developed in Section 2.5.4, the optimization model for multilayer AR coatings, and an algorithm to determine the reflectivity and transmissivity characteristics of thin-film interference filters.

1.2 The Quantum Cascade Laser

The Quantum Cascade Laser (QCL) is a semiconductor laser based on inter-subband transitions in a structure of quantum wells and barriers. In that sense its principle of operation is fundamentally different to the inter-band operation of conventional diode lasers that rely on the recombination of injected electrons and holes to emit photons. Since in the latter, the radiative transitions take place between the conduction and valence bands, the emission wavelength is determined by the width of the bandgap, which is a material property. Thus. in order to change the wavelength, a different material needs to be used.

Typical bandgaps of III-V semiconductors such as gallium arsenide, indium phosphide, and gallium antimonide are in the range of 0.7-1.4 eV and reach up to 3.4 eV in gallium nitride. Photonic energies corresponding to these transitions lie in the range spanning from the near infrared to the ultraviolet and can be modified by mixing in various percentages of different elements into the compounds, creating ternary or quaternary alloys.

To achieve emission in the MIR, e.g. 3-12 μm, with interband lasers, the bandgap needs to be between 0.1-0.4 eV, mostly only achievable with extreme mixing ratios within the lead-salt family e.g. $Pb_xSn_{1-x}Te$, and certain quarternary alloys. These material have very bad mechanical properties that severely limit their lifetime. In addition, they only work at cryogenic temperatures and produce very low power levels. Thus they do not make for good lasers.

QCLs instead use the mature technology of well-controllable wide-bandgap material systems used in high-power interband lasers. To reach long wavelengths, the transitions take place not across the bandgap, but between different, more closely spaced confined states within the conduction band only. This is achieved as follows. When a several-nanometer-thin layer of a material with a smaller bandgap is epitaxially grown between layers of a material with a wider bandgap, the Fermi levels align in such a way that both the conduction band and the valence band form abrupt band edge discontinuities at the heterojunctions. For QCLs, only the discontinuity of the conduction band is relevant for now, since they are unipolar and so far exclusively realized with electrons.

For these double heterojunctions, within the effective mass approximation, the wavefunctions of the conduction band electrons can be decomposed into the Bloch waves of the bulk materials and a slowly varying, one-dimensional envelope function across the structure, in the direction normal to the plane. For many purposes of semiconductor devices, it is useful to disregard the quickly-oscillating Bloch functions of the crystal structure and derive all the transport properties of the heterostructure by solving for the envelope function. The quantum mechanical potential, whose eigenstates the envelope functions are, is just a piecewise constant, one-dimensional potential, whose local value is determined by the conduction band edge in the Γ-point. In the scenario, where a thin layer of narrow bandgap material is sandwiched between wider-gap materials, this one-dimensional potential forms a *quantum well*, which is why this exact double heterojunction is simply called a "quantum well". To solve for the bound eigenstates of the structure – the *subbands* – the envelope function and its

derivative are simply matched at the interfaces, with the exception that the different effective masses of the materials have to be appropriately considered.

QCLs use the transitions between these different subbands of a quantum well within the conduction band to reach long wavelengths. Figure 1.3 shows the different cases for interband and inter-subband transitions. The left figure shows a crude sketch of a radiative interband transition, in this case for a quantum well both in the conduction band and the valence band. Here, the quantum well only serves the purpose of spatially confining and thus concentrating the injected electrons and holes in a very small volume, thus increasing the probability for recombination. The right figure shows the quantum well of the conduction band only, with the energy levels of two subbands and a radiative transition between the two drawn inside.

The energetic positions of the subbands can now be adjusted by simply altering the width of the quantum well within the same material system, because the band offset between the well and barrier materials is a constant. Along with the positions of the subbands, their spacing within the same well can be altered as well. This way, emission wavelengths up to the far infrared and Terahertz ranges can be achieved.

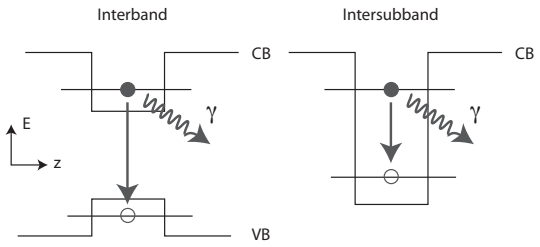


Figure 1.3 *Left: Sketch of a radiative interband transition for a quantum well (formed by the band edge discontinuities at the heterojunctions) both in the conduction band and the valence band. Here, the quantum well only serves the purpose of spatially confining and thus concentrating the injected electrons and holes in a very small volume, thus increasing the probability for recombination. Right: Radiative transition between two subbands of a quantum well of the conduction band as happening in a QCL.*

The basic principle of the QCL as originally proposed by Kazarinov and Suris in 1971 [9] is as follows. An electron is injected into an excited state, or subband, of a quantum well by applying an electric field to the structure and running a current through it. This electron then relaxes to a lower subband and emits a photon in the process with an energy equal to the spacing of the subbands. The electron is then quickly extracted from the lower state, occupying it for a much shorter time than the excited state. This process is cascaded and the electron is injected into the equivalent excited state of the next quantum well downstream through resonant tunneling. Through the different life times of the electrons in the upper and the lower subbands, a *population inversion* is achieved, which is the requirement for laser action as discussed further in Section 1.4. This process is sketched in Fig. 1.4.

For the QCL's first experimental demonstration in 1994 [10], the group of Federico Capasso at Bell Labs had to develop a number of strategies to overcome the practical challenges involved with realizing the original concept. First, they inserted a thin barrier into the quantum well that contains the laser transition, thus effectively creating two separate wells

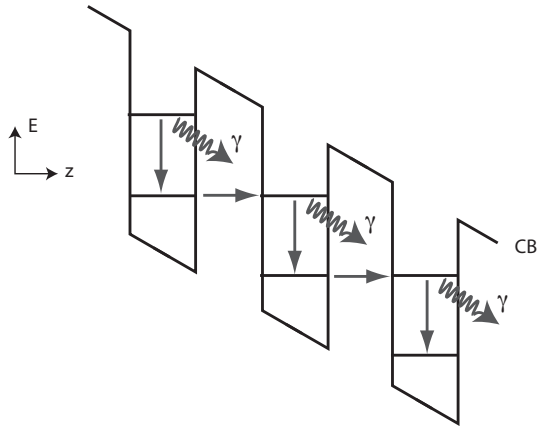


Figure 1.4 *Basic principle of operation of a QCL. An electron is injected into an excited subband of a quantum well. This electron relaxes to a lower subband and emits a photon in the process with an energy equal to the spacing of the subbands. The electron is then quickly extracted from the lower state by resonantly tunneling into the excited state of the next period, creating a cascaded process.*

between which the laser transition takes place. This is called a *diagonal transition*. Through this step, they reduced the spatial overlap between the wavefunctions of the upper and the lower state, reducing the transition matrix element and thereby increasing the lifetime of the upper state. This in turn increases the population inversion. The second effect of this step is the ability to adjust the positions of the upper and the lower laser level independently through the widths of their respective wells.

Second, they added another, a third well with a similar width as the second, separated again by a very thin barrier, creating another, lower, state with strong overlap and consequently strong coupling to the lower laser level but with an energy separation equal to the the energy of a longitudinal optical (LO) phonon of the crystal. As a consequence, electrons in the lower laser level scatter resonantly with LO phonons, allowing an ultra-fast transition to the lowest level. This process depopulates the lower laser level very quickly, which again increases the population inversion. These three quantum wells and their respective barriers are called the *active region* of the quantum well structure. Figure 1.5 shows a sketch of a diagonal laser transition within a QCL in a position-energy diagram. Level 3 is the upper laser level, Level 2 is the lower laser level and Level 1 is the "ground state" of the current active region. τ_{32} is the inverse of the scattering rate of electrons from level 3 to 2, which, if there are no other channels for depopulation, is equivalent to the life time of the upper laser level, τ_{21} is the same for the lower laser level, thus to achieve population inversion, the relation $\tau_{32} \gg \tau_{21}$ must hold.

Contrarily to the true one-dimensional case, these subbands are not actually discrete states, since in the plane of the layers, the carriers are free to move, resulting in a quasi-continuous set of states with parabolic dispersion as a function of the in-plane k-vector, $k_{||}$.

Figure 1.6 shows a sketch of a diagonal laser transition within a QCL in a mixed position-momentum-energy diagram. Figs. a), b), and c) show the (near) parabolic dispersions as a function of $k_{||}$ of the relevant subbands in the three quantum wells forming the active region. Subband 3 denotes the upper laser level, Subband 2 the lower, and Subband 1 is the ground

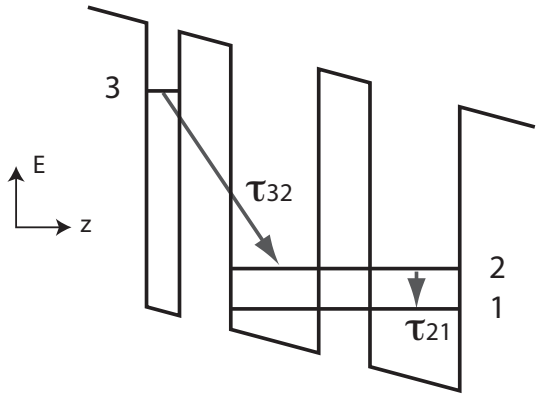


Figure 1.5 *Sketch of a diagonal laser transition within a QCL in a position-energy diagram.*

state of the period. Horizontal arrows denote tunneling processes that preserve energy and momentum. Vertical arrows preserve momentum. First, the electron tunnels from a state in Subband 3 in well a) into the virtual state in Subband 3' in well b), from which it radiatively relaxes into Subband 2 (black vertical arrow) emitting a photon of energy $\hbar\omega$. From there, it tunnels into well c) (the state overlaps with well c)) and relaxes into Subband 1 by emitting an LO phonon of energy $\hbar\Omega$ (vertical arrow). Since the dispersions of Subbands 1 and 2 are nearly parallel and have energy separation $\hbar\Omega$, electrons with any momentum $k_{||}$ in Subband 2 can scatter into Subband 1 with very small momentum transfer, making this process very fast.

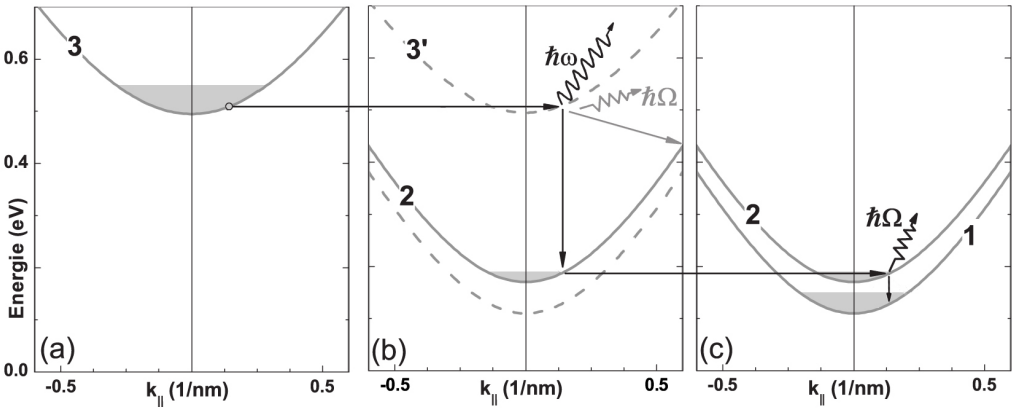


Figure 1.6 *Mixed position-momentum-energy representation of a diagonal transition in a QCL active region.*

The third development undertaken by Capasso's group was to extract the electrons from this lowest Level 1 by letting them resonantly tunnel through an extraction barrier into a region called the *injector region* (of the next cascade). The injector region is probably the most ingenious invention that was most crucially responsible for the success of the first QCL. The injector region is a *super lattice*, i.e. a structure of multiple coupled quantum wells, in which the electronic states form so-called *minibands* similar to the bands of the crystalline solid. These miniband states can be calculated similarly to the bands of the solid using the Kronig-Penney-Model. Except for the fact that the periodic potential is not formed of atomic

potentials but quantum wells, and instead of the full electronic wavefunctions, the envelope functions are derived. Since an electric field is applied to the super lattice during operation (which effectively tilts the structure in a position-energy representation), the energy levels of identical wells would shift downwards for each subsequent well, decoupling the states and localizing the wavefuntions (Wannier-Stark-Functions) in their respective wells. This effect is compensated by *digitally grading* the super lattice, i.e. reducing the width of each subsequent well while preserving the period of the lattice. This raises the energy of the subsequent wells' states to the same energetic level as the first and reinstates the good coupling of the wells. This creates fast carrier transport through the injector region, since electrons can now freely move through the miniband states while an electric field is applied. One active region and one injector region are called a *period* of the cascade (not to be confused with the period of the super lattice within the injector region).

The function of the injector region becomes clear from the following. Since the subbands have a near parabolic dispersion in the k-vector component parallel to the layers (only broken by the non-parabolicity due to the finite validity of the effective mass approximation), and as such, the dispersions of different subbands are parallel to each other, there is no forbidden energy gap between the upper and the lower laser level (see Fig. 1.6 b)). This is in great contrast to interband lasers, where there are no allowed states between the conduction and the valence band. Consequently, through interaction with phonons, electrons can relax from the upper to the lower laser level non-radiatively (grey diagonal arrow in Fig. 1.6 b)). This process is very fast and is the limiting factor in the lifetime of the upper level. Thus, electrons in the upper laser level have a radiative and a non-radiative relaxation channel towards the lower laser level, and with a certain probability, each electron will take one of the channels. The electrons that relax non-radiatively though, gain a large in-plane momentum and with it, a large kinetic energy. This is the energy they would have given off in the form of a photon if the transition had been radiative. Without the injector region, these electrons would escape into the continuum that lies above the top of the barriers of the next period and be lost from the process instead of tunneling resonantly into the excited state of the next period's active region. But transport through the injector region with its multiple miniband states gives these electrons enough time to relax to the bottom of the miniband again through emission of multiple LO phonons. Thus it cooles the electrons down and "funnels" them into the upper laser level of the next active region, again through resonant tunneling through an injection barrier. This is where the injector region derives its name from. By collecting the non-radiatively scattered electrons, the injector region serves to create periodic conditions throughout the cascade, which is a requirement for the structure to be cascaded at all. The ability to cascade this process with an essentially arbitrary number of cascades, typically a few tens to a few hundred, is the great advantage of QCLs, allowing them to achieve very high quantum efficiencies through "recycling" the electrons. This again stands in contrast to interband lasers, where after the recombination with a hole, the electron is lost from the process and needs to be replaced by freshly injected carriers.

The injector region has another function. To avoid electric field domains, i.e. an inhomo-

geneous distribution of the electric field through the QCL, the entire structure needs to be electrically neutral. This is not the case if the charge carriers are ballistically injected, so the structure needs to be doped. But if the ionized donors were in the active region, they would serve as scattering centers for the electrons in the upper laser level, thereby reducing the lifetime. By doping the center of the injector region, i.e. spatially separating the ionized donors from the active region, this effect is avoided.

Figure 1.7 shows the full conduction band edge structure of an exemplary QCL (HU2-0324, used often throughout this thesis) and the probability density of the relevant wavefunctions plotted into it (their vertical position corresponds to their energy level). The thick black lines show the upper and lower laser levels with a transition energy of 280 meV, which corresponds to a wavelength of 4.4 μm and a wavenumber of 2260 cm^{-1}. The quantum wells to the left and the right of the transition well corresponds to wells a) and c) in Fig. 1.6. To the left and right of these lie the injector regions of the current and the next period of the cascade, drawn along with their densely lying miniband states (gray). The energy separation to the next (higher lying) miniband is also drawn into the diagram with a value of 80 meV. The higher this gap is, the lower the probability for thermal carrier escape.

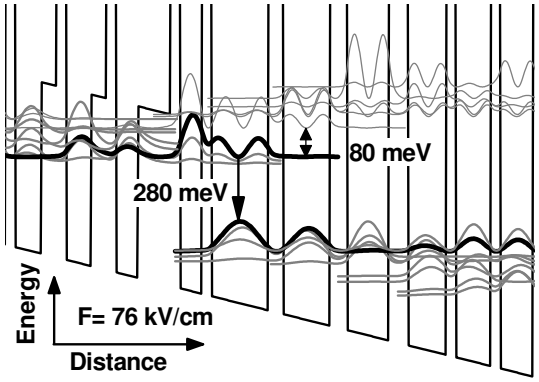


Figure 1.7 *Conduction band diagram and probability functions calculated within a single period of the active region with a 76-kV/cm electric field of an exemplary QCL (HU2-0324). The layer thicknesses in nm from left to right starting from the injection barrier (thickest composite barrier) are: 3.0/**0.9**/2.0/**0.9**/5.2/**1.4**/4.3/**1.6**/3.8/**1.8** /3.5/**0.9**/3.2/**0.9**/3.0/**0.9**/2.4/1.4/**0.9**/2.3/1.4/**0.9**/2.0. AlAs layers are in bold, $In_{0.73}Ga_{0.27}As$ layers are in roman, and $In_{0.55}Al_{0.45}As$ layers are in italics. Underlined layers are doped to 5×10^{17} cm^{-3}. The moduli square of the wavefunctions responsible for the laser transition are drawn with the thick black lines. Ground injector/extractor states are drawn with the thick gray lines and the lowest excites states are drawn with the thin gray lines.*

To avoid escape of the carriers into the continuum over the top of the barriers, the largest feasible energy separation between the upper and lower laser levels is approximately $\Delta E_C/2$, where ΔE_C is the conduction band offset between wells and barriers. But ΔE_C is a constant for a given material system. Lasers for this thesis were grown in the material system InGaAs-InAlAs on InP substrates, where InGaAs is the well material and InAlAs is the barrier material. If the relative concentrations are such that this material system is lattice matched to InP, i.e. $In_{0.53}Ga_{0.47}As$ wells and $In_{0.52}Al_{0.48}As$ barriers, the conduction band offset is $\Delta E_C = 530$ meV, which results in a shortest possible wavelength of approximately 4.7 μm.

An approach pioneered by Semtsiv et al. [11, 12] uses pure AlAs as barrier material to create barriers approximately three times as high, allowing for significantly shorter wavelength QCLs. This material experiences high levels of tensile strain from the substrate, which has to be carefully compensated for by using compressively strained well material with the composition $In_{0.7}Ga_{0.3}As$. Since every period of the QCL now has to be strain-compensated, however, there is one degree of freedom less in the design. This freedom is reintroduced by creating composite barriers using a variable amount of $In_{0.55}Al_{0.45}As$, which is slightly compressively strained against InP. This approach is termed *strain-compensation with composite barriers* and the resulting barrier height is raised from $\Delta E_C = 530$ meV to $\Delta E_C = 1400$ meV. This can be seen in the steps of the three left-most barriers plotted in Fig. 1.7. These QCLs are grown in our labs using gas-source molecular beam epitaxy (GS-MBE) to accurately control the very high strain levels in addition to the ability of growing high-purity InP for buffer layers and waveguide claddings.

Emission from a QCL is plane polarized with an electric field vector parallel to the growth direction. This can be understood from the following argument. The operator for dipole transitions is proportional to the position vector $\vec{r}$, which has odd parity. For the matrix element of the operator not to vanish, the initial and final states need to be of opposing parity. But the wavefunctions of the subbands are a product of the Bloch states and the envelope function. In interband lasers, the electronic transition is between the conduction and the valence bands, whereas the Bloch states of the conduction band are s-like and the ones of the valence band are p-like. Since s-like bands have even parity and p-like bands have odd parity, all transitions are generally allowed. Inter-subband transitions in a QCL, however, all take place within the conduction band, so the parity criterion must be fulfilled for the envelope function, which is only the case for transitions between subbands of opposing parity. Since the subbands with odd principal quantum number, $n = 1, 3, ...$, are even and subbands with even principal quantum number, $n = 2, 4, ...$, are odd, only odd to even or even to odd transitions are allowed. However, for the in-plane component the selection rule cannot be fulfilled, since for this component the wavefunctions are free Bloch states without (or a constant) envelope function. Thus, if a radiative transition does takes place, the emitted photon must have an electric field vector that is parallel to the growth direction.

After the quantum well structure is grown, the wafers are lithographically processed. Ridges, so-called *mesas* are etched to confine the pump current flow laterally to a thin stripe. These structures are then laterally electrically passivated. This is done using a transparent insulator such as silicon dioxide that doubles up as a waveguide cladding to better confine the light field to the active region (from now on in this thesis, the term “active region” denotes the ridge of the quantum well structure including all the periods of (optically passive) injector regions and laser transition regions that were previously termed the “active region”). The InP substrate and an InP layer grown on top of the structure serve as the top and bottom waveguide cladding.

Due to the non-radiative recombination channels and the consequently large current flow necessary to achieve population inversion, a QCL produces a large amount of heat. This

heat is highly detrimental to the performance since it increases the probability for non-radiative relaxation as well as facilitating thermal backfilling of the lower laser level and thermal escape into the continuum. Especially in order to drive the QCL in continuous wave mode – i.e. not with discrete current pulses, between which the laser has time to cool off – an exceptionally good heat extraction out of the active region is imperative. Since silicon dioxide has a very low thermal conductivity, modern QCLs are not passivated using SiO_2, but instead are epitaxially overgrown using InP, which has a very high thermal conductance. In order to fabricate electrically insulating InP, it is doped with iron, which serves as deep acceptors for the residual carriers. This structure is termed a *buried heterostructure*.

After the area above the ridge is opened up electrically in a lift-off process, the structure as well as the bottom of the substrate is gold-plated to create the Ohmic contacts. The wafer is then cleaved in two planes normal to the ridge waveguide to create very smooth waveguide ends, called the *facets*, and cleaved again to isolate a single ridge. The QCL is then soldered onto a copper heat sink and the top is electrically contacted using wire-bonding techniques. In order to further increase the heat extraction, instead of wire bonding the QCL, it is often soldered upside-down onto a thin gold-coated aluminum nitride submount, which has the advantage of high thermal conductivity paired with a high electrical resistivity. The submount is then soldered onto the copper heat sink and the gold layer facing the QCL and the upward-facing bottom of the substrate are then electrically contacted for the current to flow. This technique is termed *epi-down* mounting.

Over the years, many different active region designs have been developed to increase the overall performance – i.e. low threshold current density, high power, broad gain, and high temperature stability of both the threshold current density and the slope efficiency – as well as reaching different wavelength regions. Although the first QCLs have already reached a very high level of maturity and sophistication, research on this topic is ongoing and vivid. Ample theory on QCLs is available in dedicated textbooks, such as [13–15], and reviews such as [16]. Historical overviews are available, among others, in [17] and [18].

1.3 Brief Historical Overview of External Cavity QCLs

The first experimental demonstration of an External Cavity mid-infrared semiconductor laser was undertaken in a Littrow configuration by Han Q. Le et al. from the Masachussetts Institute of Technology in 1996 [19]. This was followed by an in-depth analysis by the same group in 1997 [20]. Although the first QCL had already been demonstrated two years before, availability was still limited, so the first demonstration of an EC MIR semiconductor laser was done using Sb-based, optically pumped interband lasers at cryogenic temperatures.

However, the experimental procedure and general analysis of a Littrow setup for semiconductor MIR lasers remained highly relevant to the first (Littrow) EC-QCLs. This was probably the reason why the first experimental demonstration in 2001 was undertaken by associates of Mr. Le's, by now at the University of Texas in Houston, Guipeng Luo and Chuan Peng [21]. The in-depth analysis of this setup, written a year later [22], was therefore

largely taken over from the paper in 1997.

The first room-temperature single-mode emission from an EC-QCL [23], now at 10.4 μm instead of the 4.5-5.1 μm range before, was undertaken by Totschnig et al. from the University of Vienna in 2002, with a QCL from Alpes Lasers founded by one of the inventors of the QCL in Capasso's group at Bell Labs, Jérôme Faist. For the next few years, both Mr. Le's group [24, 25] as well as Richard Maulini [26, 27] from Mr. Faist's group, now at the University of Neuchâtel, were actively working on EC-QCLs, quite exclusively in the Littrow configuration. As soon as 2002, the first antireflection-coated QCLs were used. By 2003, Mr. Le's group has also turned to acquiring QCLs from Alpes Lasers, after the first works had used their proprietory QCLs.

In 2005, Gerard Wysocki from Frank Tittel's group at Rice University has joined forces with Mr. Faist's group, to make the first mode-hop free EC lasers [28] – by mounting the grating on a piezo-controlled actuator to actively control the cavity length – based on an EC version earlier developed by Mr. Maulini, for chemical sensing applications [29].

By 2005, just four years after the first demonstration of an EC-QCL, a commercial company, Daylight Solutions, was founded by Timothy Day in Poway, CA, USA, to develop and sell miniaturized Littrow QCLs. For this, the co-founder of New Focus adapted the Littrow technology of External Cavity diode lasers used in VIS-NIR telecommunication applications for the use with mid-infrared QCLs.

One year later, another company, Pranalytica Inc. founded by the inventor of the CO_2 laser, Kumar Patel, in Santa Monica, CA, USA, has also started developing commercial EC-QCLs in the Littrow and Littman-Metcalf designs, originally using QCLs from Manijeh Razeghi's group at Northwestern University. Aside from the tunable lasers, Pranalytica also builds full EC-QCL-based gas sensors for rugged environments [30, 31].

Since then, both of these companies have performed extensive development of their product range, now even including modular laser systems in compact housings using multiple QCLs spanning a wide tuning range (Daylight's MIRcat (TM) and Pranalytica's OmniLux (TM) product lines). They also continuously update their product lines with state-of-the-art QCLs. While Pranalytica invests considerable development efforts into QCLs, Daylight has a range of partially undisclosed suppliers. However, they never left their core (Littrow) technologies [32–36].

Within a short time, these two companies became the main suppliers of EC-QCLs for most academic spectroscopic MIR applications, nearly exclusively carrying the research and development efforts on Littrow EC-QCLs worldwide. While spectroscopic applications utilizing EC-QCLs purchased from these companies have multiplied considerably in recent years, only several groups have continued to develop and analyze proprietary External Cavities for different applications.

These include the following. Gensty et al. have investigated and characterized intensity noise in 2005 [37], Mohan et al. reported continuous wave (CW) operation at room temperature in 2007 [38], Mukherjee et al. have characterized the emission linewidth using NO_2 saturation spectroscopy in 2008 [39]. Maulini et al. demonstrated a widely tunable CW

EC-QCL at 4.6 μm in 2009 [40], Hinkov et al. have performed time-resolved spectral characterization the same year [41]. Knabe et al. have performed frequency metrology in 2012 [42], Leonhaeuser-Rein et al. have investigated mode beat frequencies in 2013 [43]. The same year, Inoue et al. have characterized the intensity fluctuations [44], and Tan et al. have presented room-temperature CW at 4.7 μm with a very low threshold current density [45].

Most noticeably, the Wysocky group at Princeton University, the Fraunhofer Institute for Applied Solid State Physics (Germany), and the Optics and Infrared Sensing group at Pacific Northwest National Laboratory continue to use proprietary EC-QCLs for spectroscopic applications.

The Wysocky group continues to use their setup based on the Maulini design. This setup has, among others, been used by Weidman et al. from Oxford University to investigate heterodyne spectro-radiometry [46, 47]. The Fraunhofer Institute also produces their own QCLs. Here, Fuchs [48, 49] and Hinkov [50] have performed Imaging Standoff detection of explosives. Phillips et al. from Pacific Northwest National Lab use QCLs acquired from Maxion, Inc., and incorporated it into a Littrow cavity for trace gas detection using a Quartz tuning fork (QEPAS, explained shortly) [51] and a multi-pass Herriott cell [52], and intracavity absorption [53]. They have also performed hyperspectral imaging [54]. Finally, Stupar et al. [55] and Sonnabend et al. [56] from the Universität of Köln evaluated fully reflective Littrow Cavities for heterodyne spectroscopy.

The number of publications on spectroscopic applications using EC-QCLs from Daylight Solutions is far greater. The group of Frank Tittel at Rice University is very active in the field of Quartz-enhanced photo-acoustic (QEPAS) spectroscopy. In photoacoustic gas spectroscopy (PAS), the analyte is periodically illuminated with a laser beam modulated at at acoustic frequencies. If the wavelength of the light coincides with an absorption line, the molecules are excited, which results in periodic heating and expansion, which can be picked up by a microphone that produces a signal proportional to the strength of the absorption. In QEPAS, the analyte is located between the branches of a tuning fork, whose resonance frequency is used to modulate the laser beam. This increases the sensitivity by orders of magnitude compared with conventional photoacoustic spectroscopy. The signal is then picked up by a piezo-electric Quartz located at the joint of the branches, producing an AC voltage. Lewicki, Kosterev, Gong, Dong, and Kachanov have demonstrated the technique on a number of different molecules, such as Freon 125 [57, 58], NO [59, 60], NH_3 [61], CO [62], water vapor, aceton, and ethanol [58].

Vincenzo Spagnolo and Pietro Patimisco from the University and Politecnico of Bari, Italy, have collaborated with the Tittel group and created hollow fiber-coupled QEPAS sensors, also using Lasers from Daylight solutions [63–66]. Hollow waveguides are also used by Brandstetter et al. for biomedical analyses of glucose and lactate in blood plasma and other biofluids [67–70].

Young et al. from the Georgia Institute of Technology, have collaborated with Daylight Solutions on hollow waveguide sensors to detect multianalytes of ethyl chloride, dichloromethane, and trichloromethane [71]. Hancock, Walker, and van Helden from the University of Oxford

have created and analyzed effects in sensors for NO, water vapor and Deuterium Bromide [72–74]. Karpf and Rao from Adelphi University in Garden City, New York, used Daylight Lasers and multi-line analysis [75–77] and in conjunction with high-finesse optical cavities in off-axis [78] and cavity-ringdown [79] configurations. Chao et al. from Stanford University have performed realtime in-situ analysis of NO [80, 81].

Kottmann et al. from the ETH Zurich and Pleitez et al. from the Goethe-Universität Frankfurt used a Daylight Laser to detect glucose in the human epidermis using photoacoustic spectroscopy [82–84]. Wen et al. from CANMET Energy Technology Centre in Devon, Canada, performed photoacoustic spectroscopy on solids [85]. Also in Canada, Sydoryk and Parsons at the University of Alberta used a Daylight Laser with a Harriet cell to monitor benzene and toluene simultaneously and in realtime [86, 87]. Woerle et al. from the University of Ulm in Germany perfomed breath analysis on mice to detect the $C^{13}O_2/C^{12}O_2$ isotopologue ratio [88]. Suter et al. from Pacific Northwest National Laboratory used a Daylight laser for standoff detection of explosives [89]. Knabe et al. from the National Institute of Standards and Technology in Boulder, Colorado, collaborated with Daylight to detect N_2O with a comb-calibrated EC-QCL [90]. Kole et al. from the University of Illinois at Urbana-Champaign, have performed microspectroscopy and imaging [91]. Luedeke et al. from the University of Freiburg in Germany used a Daylight laser to examine vibrational circular dichroism in three chiral compounds [92]. Ribaudo et al. from the University of Massachusetts Lowell investigated plasmonic gratings using an EC-QCL [93].

One of the main focuses of Pranalytica are security and defense applications. Michael Pushkarsky and Michael Webber, among others, have published highly sensitive detection of TNT [94] and sub-ppb-level NO_2 [30] using Pranalytica's EC-QCLs. Mukherjee et al. from the University of California have collaborated with Pranalytica to detect sub-ppb-levels of DMMP (Dimethyl methylphosphonate), a chemical warfare simulant [95] and to resolve the fine structure of NO_2 [96]. Pranalytica also produces trace gas detection systems in conjunction with photoacoustic cells [31].

While this list is just nearly comprehensive, it gives a good overview of the general activity in the field. Since the suitability for spectroscopic applications is one of the main features of infrared QCLs, optimizing their performance in these has become one of the main objectives for QCL research and development. For this reason, one of the main figures of merit of QCL development has become the broad and flat gain profile, which results in a wide tuning range in an EC configuration. Due to the great design flexibility of QCLs, there are very different ways to achieve this. The two most promising approaches are the following. The first, the *heterogeneous cascade* approach, stacks blocks of cascades with different design wavelengths on top of each other within the same ridge waveguide. The second, the multiple-upper and/or multiple lower state approach, sometimes termed *continuum-to-bound* or anti-crossed *dual-upper-state* designs, uses homogeneous cascades with multiple laser transitions to broaden the spectrum.

Richard Maulini pioneered the first two-stage heterogeneous cascade in 2006, which, inserted in an External Cavity, was tunable between 8.2 and 10.4 μm or 265 cm^{-1} in pulsed

mode, which corresponds to 24% of the center frequency [97]. Andreas Wittmann, also with the Faist group, by now at the Institute of Quantum Electronics at the ETH Zurich, took over this work and expanded the tuning range to 292 cm^{-1} or 25% in 2008 [98]. Andreas Hugi, from the same group, then proceeded to stack cascades emitting at 5 different wavelengths in a symmetric manner in 2009 [6]. Here, the shorter wavelength cascade block was placed in the center of the waveguide and the longer wavelength cascade blocks symmetrically around it. Since for shorter wavelengths the light confinement is stronger, this strategy reduces overlap of the higher-energy photons with the outer cascade blocks, which in turn reduces cross-absorption. This structure was tunable from 7.6 to 11.4 μm, or 432 cm^{-1}. Sabine Riedi, also from the Faist group, built on the work of Alfredo Bismuto [99] to achieve 556 cm^{-1} in the 3-4 μm range using two cascades emitting at 3.3 and 3.7 μm [100] in 2013. To reach this short wavelength range, they used the strain-compensated composite barrier approach developed by Mykhailo Semtsiv et al. in 2004 [11, 12].

The second approach to broaden the tuning range is pursued successfully, among others, by the group of Claire Gmachl, a former colleague of Mr. Faist's in Federico Capasso's group at Bell Labs, at Princeton University. Yu Yao developed a homogeneous design with three upper states, tunable by 200 cm^{-1}, from 7 to 9 μm in 2010 [7]. The designs of this approach are more complex than the designs of each period in the heterogeneous cascade approach. However, the necessity to optimize the doping concentration of each cascade and the electron transfer between the cascades, as well as the arrangement of the cascades within the waveguide to avoid spatial holeburning and gaps within the tuning range is avoided. Also, the stability with respect to the pump current is increased. In 2011, Tatsuo Dougakiuchi from Hamamatsu Photonics K.K. designed an active region with two upper and multiple lower states that is tunable over 321 cm^{-1}, or 21% around 6.8 μm [101].

Several different External Cavity QCL configurations have been demonstrated over time, however, their number and diversity is far smaller than in the VIS-NIR range, as reviewed for instance by Mroziewicz [102]. Aside from the fact that diode lasers in this range have matured decades earlier than QCLs, this is also owed to the ease of availability of various different optical components in the VIS-NIR range, such as well-working acousto-optic or electro-optic filters, that have not yet been developed with similar performance in the MIR. Consequently, the vast majority of EC-QCLs achieves broad tuning using a simple diffraction grating.

The simplest and most successful setup, as mentioned earlier, is the Littrow setup, which exists in four different sub-designs, mainly differing by whether the useful beam is coupled out of the extra-cavity facet (the facet not facing the grating) or whether this facet is not accessible, for instance because it is coated with a high-reflectivity (HR) coating. Fig. 1.8 shows the four designs. Design a), the most wide-spread design, as used for instance by Daylight Solutions, couples the beam out of the extra-cavity facet. The great advantages of this design is large output power emitted by a long laser stripe in pulsed mode, as well as the fact that the output beam direction does not shift while tuning the laser. Configuration b) uses an HR-coated back facet, which is best used in conjunction with short laser stripes driven at low driving currents in CW, where the coating is necessary to reduce the losses of

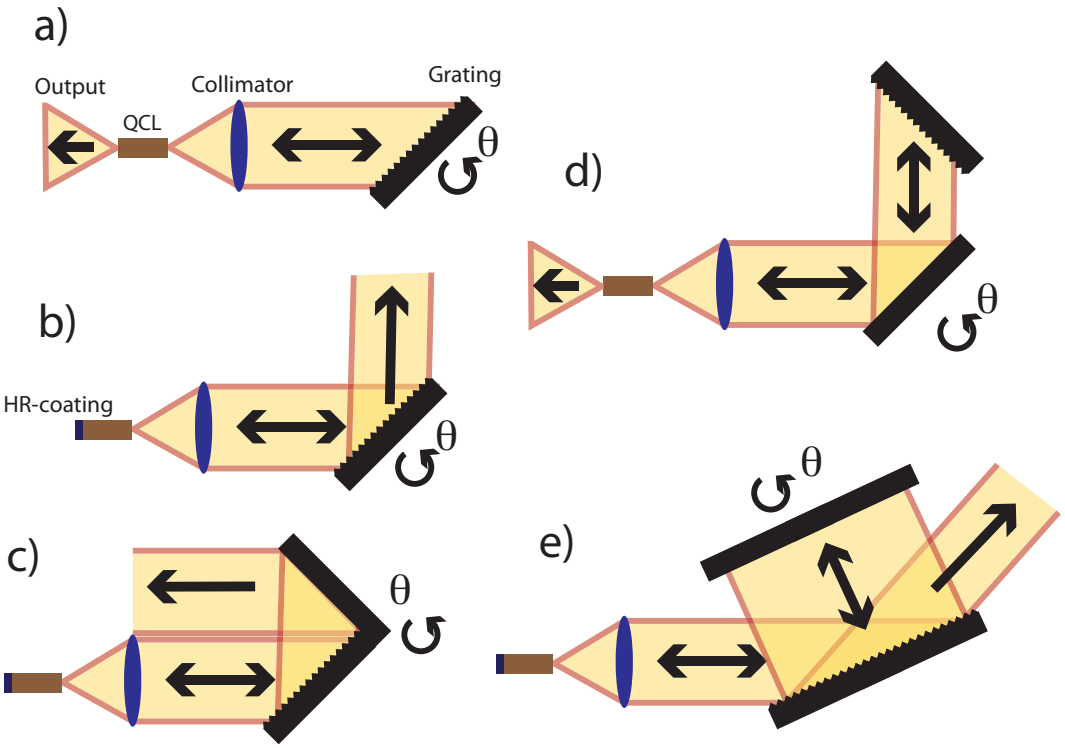


Figure 1.8 *a)-d): Different configurations of the Littrow design. e) Littmann-Metcalf design*

the QCL to sustain the lasing condition. The useful beam is now collected from the zeroth-order reflection off the diffraction grating. This design has two major drawbacks. The first is that there is normally very little intensity at the zeroth order, so little useful light, because the gratings are designed to concentrate their efficiency on the first order reflection used for the Littrow feedback. But this, of course, is by design, to allow for little losses out of the cavity. The second drawback is that as the grating is tuned by an angle θ, the beam walks off by an angle 2θ. This is solved in setup c), pioneered by Maulini [26, 27], where a mirror, aligned at right angle with respect to the grating reflects the beam back parallel to itself. By choosing the pivot point of grating and mirror correctly, a lateral (parallel) walk-off can also be eliminated. Design d) uses a folded Littrow cavity (pioneered by Zhang and Tsai from the Wysocki group) [103–105]), where the beam is deflected by an angle-tunable mirror onto a static grating. This mirror can either be piezo-actuated or quickly rotated. Either way, this setup tends to tune much faster than the other three, since a mirror can generally have much smaller mass than a grating. Designs a)-c) can be made mode-hop free by actively altering the (effective) cavity length (e.g. Wysocki's design [28]). This can be done either by mounting the grating on a piezo actuator, or by altering the effective optical length of the QCL chip by thermally varying its refractive index. One particularly crafty way, although mechanically highly fragile, is to choose the pivot point of the grating in such a way that as the grating is rotated, the cavity length is varied along automatically so that the order of the single resonator mode stays constant. The laser mode then "surfs" on the maximum of the Littrow reflection [106, 107].

The Littmann-Metcalf design, configuration e) in Fig. 1.8, uses a static grating that dis-

perses the light in first order in the direction of an angle-tunable mirror. The mirror selects the wavelength with normal incidence by reflecting it back on the same path, thus being thrown back into the QCL after a second reflection from the grating. The useful light is then collected from the zeroth order reflection off the grating. The main advantage of this setup is a narrower bandwidth, since the grating filters the wavelength twice. The main drawback is an added component that requires fine alignment.

There have been demonstrations of MEMS-scale gratings [108], however, while their usefulness may be justified in the visible and UV ranges, it is highly questionable in the MIR. This is because, as discussed in Sections 1.5.2 and 5.5.3, the bandwidth of the light filtered by the diffraction grating is proportional to the number of illuminated grooves. But the groove distance has a lower limit, half of the wavelength to be reflected in first order, so smaller gratings will necessarily have less selectivity leading to multimode emission. This is discussed in Chapter 5 along with our contribution to solve the problem of miniaturizability.

Another very interesting concept is the external ring cavity pioneered by Malara et al. [109]. This setup does not use a wavelength-selective element, instead, spectroscopy can be performed using intra-cavity absorption, where the mode-competition in the transient phase of the multimode laser oscillation is observed over time in the presence of an analyte within the cavity. However, the precise alignment of the 8 optical components and a gas cell will pose limitations on the feasibility of handheld devices.

An interesting concept for a tunable transmissive wavelength filter for EC setups has been demonstrated by Wassermann et al. [110]. They observed surface-plasmon-enhanced extraordinary transmission through a thin metallic film with sub-wavelength perforations at a certain wavelength. The wavelength of this extraordinary transmission could be tuned by tilting the filter.

Finally, although not an External Cavity configuration, one interesting concept for QCL emission with broad tunability also needs to be mentioned. The concept, pioneered by Benjamin G. Lee in the group of Federico Capasso (one of the inventors of the QCL) at Harvard University, utilizes an array of parallel distributed feedback (DFB) QCLs [111–114]. As mentioned in Section 1.1, DFB QCLs have narrow-band tunability with the advantage of being very fast and monolithic. Lee has applied DFB gratings with incremental period to a set of parallel QCL stripes on the same wafer. This way, the overlapping tuning ranges with incremental central wavelengths of each DFB stripe span the entire gain of the QCL. Even this concept, however, has a drawback, which results from the spatial offset of each laser's emission, limiting its usefulness in this form. In order to spatially combine the beams, the monolithical nature has to be sacrificed by, again, introducing a lens and a grating [115–117].

A good review of the early days of EC-QCLs can be found, for instance, in Richard Maulini's PhD thesis. [118]. Andreas Hugi has reviewed EC-QCLs in 2010 [119]. Reviews of EC-QCLs for different sensing applications have been written by Tittel [120], Sigrist [121], Risby [122], Rao [123], Brandstetter [124], and Roepke [125]. QCL development for ECs has been reviewed by Wojcik [126].

1.4 Theoretical Introduction to Laser Operation

A laser in general consists of a light amplifying medium and a positive feedback mechanism for the amplified light. Often the feedback is achieved through incorporating two reflectors on either end of a linear amplifier. In the case of a semiconductor laser such as the QCL, the amplifier (*active region*) is a linear waveguide etched from high-refractive index semiconductor material with a cladding of lower-index material surrounding it. Feedback is inherently present once the substrate (with active region) is cleaved at two planes normal to the waveguide, which results in partial Fresnel reflection at the cleaved planes. The cleaved faces of the waveguide are called the *facets*. For typical QCL active regions of InAlAs/InGaAs with an effective refractive index of approximately 3.25 the *as-cleaved* reflectivity is

$$R = \frac{(3.25-1)^2}{(3.25+1)^2} \approx 28\%. \tag{1.1}$$

A simplified view of a laser amplifier is a picture of two energy levels that contain carriers, i.e. electrons in the case of a QCL. Emission of a photon occurs upon transition from the upper laser state to the lower laser state and the energy of the photon is equal to the energy spacing of the levels. In a QCL, like in many other lasers, there are not only radiative transitions that emit a photon, but also non-radiative transitions that emit phonons, producing heat. The radiative transitions can be spontaneous or stimulated by another photon of the same energy. In the latter case, the emitted photon has the same direction and phase as the photon that stimulated its emission, the two photons are coherent. Electrons in the lower laser state, however, can absorb photons of the same energy and transition to the upper level, thereby reducing the number of photons. The absorption happens with the same probability as the stimulated emission, thus, to achieve net light *amplification*, it is essential that at any given time there be more electrons available for stimulated emission than absorption, i.e. more electrons in the upper laser state than the lower. This is called *population inversion*, since it is a distribution that is not achievable in any thermal equilibrium, but only by selectively populating the upper energy level using an external energy source, i.e. *pumping* electrons to the upper level.

Laser action is achieved in the following way. The active region is pumped beyond the point of population inversion. Electrons from the upper state then spontaneously relax to the lower state, thereby emitting photons. These photons can stimulate the emission of more photons, and since there is population inversion, the photon density grows (exponentially). When the photons reach one of the reflectors, one part is transmitted and thus *lost* – this is the useful light emitted by the laser – the rest is reflected back into the active region and is available to stimulate the emission of more photons. If the pump rate is high enough that the number of photons lost – to absorption, transmission through the reflectors, and other loss mechanisms, such as scattering at free carriers – is smaller than the number of photons gained within one round trip, the overall photon density will continue to grow exponentially. Higher photon densities, however, saturate the gain of the active region and the internal steady state photon

density is reached when the number of photons lost out of the system during one round trip equals the number of photons gained. In this case, the energy supplied to the system through pumping is balanced out by the useful light that leaves the system plus the heat produced by non-radiative transitions from the upper to the lower state.

For practical reasons, real-world lasers do not have two levels only, but at least three, often four. The conduction band and the valence band of an interband laser act as continua of states that make the third and fourth level of a four-level laser, with the band edges acting as upper and lower laser levels. Cascaded lasers, such as the QCL, have a number of subbands for every period of the cascade, while the band edges of two (or more) distinct subbands (Subbands 2 and 3 of Section 1.2) behave as the laser states, and Subband 1 and the injector region serve as the third and fourth level.

Section 1.4.1 outlines the canonical mathematical treatment of laser resonators. This procedure, however, is only valid for facet reflectivities relatively close to 1, because in this case the differential equations governing the internal photon propagation have trivial solutions and most resulting expressions are analytical. Section 2.5.4 eliminates this condition and illustrates a more complete bottom-up model to more closely agree with the experimental conditions by making use of the full differential equations.

1.4.1 Canonical model for laser dynamics

When a laser is pumped above its threshold, there is a transient phase within the amplifier, where spontaneous emission experiences gain and distributed loss, as it travels through the amplifier. Assuming the *gain* γ is a constant in space and time (an effective resonator gain), the light propagates according to [22]

$$E(z,t) = E_0 e^{(\gamma-\alpha_w)z/2} e^{i(\beta(\nu)z-2\pi c\nu t)}, \tag{1.2}$$

where $E(z,t)$ is the position and time-dependent (complex) magnitude of the electric field vector, E_0 is its value at $t = 0$ and $z = 0$, α_w is the distributed waveguide loss due to scattering, $\beta(\nu)$ is the propagation constant, and ν is the wavenumber. There is a factor of 2 in the exponent because gain and loss are defined through the photon flux density ϕ that is proportional to E^2.

In steady state, the light wave reproduces itself exactly (in amplitude and phase, at all times) upon one round-trip. Thus, in the simplest form, the so-called *lasing condition* is

$$e^{l(\gamma-\alpha_w)} e^{i2l\beta(\nu)} r_1 r_2 = 1 \tag{1.3}$$

where r_1 and r_2 are the amplitude reflection coefficients of the reflectors. $\beta(\nu)$ is the propagation constant, which is just the magnitude of the wave-vector $\beta(\nu) = 2\pi\nu n(\nu) = \frac{2\pi n(\lambda)}{\lambda_0}$, where n is the index of refraction and λ_0 the associated wavelength in vacuum. In the case of a guided mode inside a laser resonator, where the index of refraction is both highly dispersive and mode-dependent, the functional form of the propagation constant is best given

approximately in terms of the first terms of a Taylor expansion around a central wavelength

$$\beta(\nu) = 2\pi n_{\text{eff}}(\nu_0)\nu_0 + 2\pi n_{\text{g}}(\nu_0)(\nu - \nu_0) + \frac{4\pi^2 D(\nu_0)}{2}(\nu - \nu_0)^2, \tag{1.4}$$

where $n_{\text{eff}}(\nu_0)$ is the effective index of refraction of the waveguide of the mode with central wavenumber ν_0, $n_{\text{g}}(\nu_0)$ is the group index, and $D(\nu_0)$ is the dispersion.

But relation 1.3 has to hold for both the real and imaginary parts of the left hand side, resulting in the two conditions

$$l(\gamma - \alpha_w) + \ln|r_1| + \ln|r_2| = 0 \tag{1.5}$$

and

$$2l\beta(\nu) + \arg(r_1) + \arg(r_2) = 2m\pi, \tag{1.6}$$

where m is an integer. Rewriting Eq. (1.5) results in

$$\gamma = \alpha_w + \frac{1}{l}\ln\frac{1}{|r_1|} + \frac{1}{l}\ln\frac{1}{|r_2|} \equiv \alpha_w + \alpha_{m1} + \alpha_{m2} \equiv \alpha, \tag{1.7}$$

where the mirror losses α_{m1} and α_{m2} and the overall loss α were introduced. It is clear that the right hand side is a constant (i.e. not a function of pump rate), which means during laser operation the effective round-trip resonator gain γ is "clamped" to the value of the losses, regardless of pump rate. At higher pump rates the photon density increases, as will become clear in Section 2.5.4, but the effective gain does not. Another statement of this fact is that during laser operation, the laser's gain is clamped to its value at threshold, γ_{th}, and that this value is equal to the loss:

$$\gamma = \gamma_{\text{th}} = \alpha \tag{1.8}$$

Gain clamping is an effect of amplifier saturation which is discussed in detail in Sections 1.5.5 and 2.5.4.

Rewriting Eq. (1.6) for the simple case of no dispersion and only (phase-independent) Fresnel reflections, so $\arg(r_i) = 0$, results in

$$\nu_m = \frac{m}{2n_{\text{eff}}l}. \tag{1.9}$$

This makes clear that there are a set of *allowed* modes ν_m that depend on the length of the resonator and its effective refractive index. These are the set of wavelengths for which the resonator can sustain standing waves.

1.5 Theoretical Introduction to QCLs in a Littrow External Cavity

For dispersive spectroscopical applications, it is a requirement to have an emission wavelength that is a) highly narrowband, best if it is single-mode, b) has a pre-determined wavelength, and c) this wavelength can be adjusted or tuned in a controlled manner during operation.

Fabry-Perot lasers, as discussed in the previous section, consist of an active region with a gain that spans over a certain spectral range and mirrors with a spectral response that is more or less flat over a typical gain. The threshold for laser action is the pump current required to raise the maximum of the gain spectrum over the value of the losses. The first mode to start oscillation is therefore located at the spectral position of the gain maximum. If the pump current is further increased to reach higher output powers, an increasingly larger part of the (unsaturated) gain curve is raised over the loss value. If the gain is strictly homogeneously broadened (all laser transitions have the same energy with lifetime-determined broadening), many modes start to oscillate when the pumping first starts, but the strongest mode sooner or later defeats the competing modes, because it experiences the strongest positive feedback and claims all the available pump energy for itself. However, due to hole-burning and temperature effects, the strongest mode tends to fluctuate in intensity and spectral position, leading to noise and increased bandwidth. In any case, lasing takes place near the maximum of the gain curve, regardless of the overall width of the gain and the output is chaotic.

An External Cavity (EC) configuration has the following purpose. It introduces a strongly wavelength-dependent loss to the laser resonator, whose value is low in a very narrow band that can be freely positioned, and very high everywhere else. This is achieved by *replacing* one of the flat-response reflectors (mirrors) with a reflector configuration that has a high reflectivity in only this band and essentially causes an *open resonator* with no reflectivity everywhere else. This way, the active region can be pumped strongly without laser action to occur except in the band where the loss is lowered, forcing the emission wavelength into this band. If the band is narrow enough to discriminate between resonator modes, a single tunable mode is selected this way. The ratio between the grating-selected on-resonance loss and the off-resonance loss determines the highest possible pump rate for exclusive EC operation. It also determines the maximum tuning range since the gain tails off towards its sides and (depending on pump rate) eventually drops below the loss value of the EC resonance.

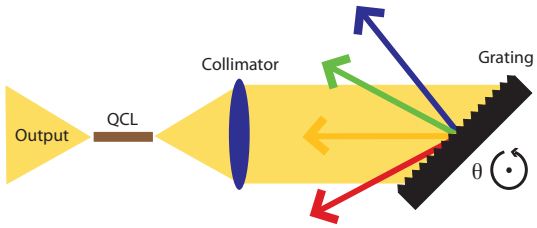


Figure 1.9 *Sketch of a Littrow-type external cavity.*

As mentioned before, the most common External Cavity configuration is the so-called

Littrow configuration, where the narrow-band reflector is a reflective diffraction grating used in first order. A sketch of the configuration is given again, for convenience, in Fig. 1.9. If a semiconductor laser amplifier such as a QCL is used, the output beam is highly divergent and has to be collimated, as is done with a lens in this sketch. The collimated beam then impinges on the diffraction grating, which disperses the light according to wavelength. One narrow band however is aimed straight back collinear to the impinging beam and is focussed back into the ("intra-cavity") laser facet through the same lens. This band is centered at the *Littrow wavenumber* given by the *Grating Equation*

$$\nu_g(\theta) = (2d\sin(\theta))^{-1}, \tag{1.10}$$

where the tuning angle θ is the angle between the incoming beam and the grating normal and d is the groove spacing.

In order to truly replace one of the mirrors of a semiconductor laser, it is not sufficient to merely add the external optics, since the cleaved facet of the gain chip inherently has a relatively high Fresnel reflectivity. This leads to coupled-cavity effects that are in most cases highly detrimental to laser operation as discussed in Section 1.5.6. Instead, the reflectivity of the facet needs to be reduced to very low values, which is normally done by using an anti-reflective (AR) interference coating. This is discussed in more detail in Chapter 2.

This section introduces the mathematical framework to analyze the effects of a QCL in conjunction with a Littrow External Cavity. Further aspects of EC operation are analyzed in detail throughout the rest of this thesis.

For QCLs, the threshold current density is normally written as [16]

$$J_{th} = \frac{\alpha}{\Gamma g} \tag{1.11}$$

where α is a sum of the losses, Γ is the confinement factor so that the effective losses are α/Γ, and g is the gain coefficient. This is a consequence of Eq. (1.8) and implies that the threshold gain is given by

$$\gamma_{th} = J_{th} g \tag{1.12}$$

and is the peak value of the gain curve of a laser pumped at threshold.

However, this is only true in the special case where α is (approximately) a constant over the entire gain, as it is for a Fabry-Perot Laser, where $\alpha_{FP} = \alpha_w + \alpha_{m1} + \alpha_{m2}$, because the Fresnel reflectivities of the laser facets that enter the mirror losses are nearly constant.

In an external cavity laser, the loss is not a constant but is a function of wavenumber, $\alpha = \alpha(\nu)$, with a sharp dip that is orders of magnitude narrower than the gain width and freely positionable. To characterize the laser output while tuning the grating feedback across the gain spectrum, both the gain and loss have to be considered in their full functional forms including their wavelength dependences, $g(\nu)$ and $\alpha(\nu)$, which are discussed in Sections 1.5.1 and 1.5.2.

Including the wavelength dependence into Eq. (1.11) leads to wavelength-dependent threshold,

$$J_{th}(\nu) = \frac{\alpha(\nu)}{\Gamma g(\nu)}, \tag{1.13}$$

whose interpretation is key to understanding many EC phenomena.

A wavenumber-dependent threshold is to be interpreted as the current density required to achieve lasing at a certain wavenumber (i.e. mode). But then, the global threshold current density of the laser is just the global minimum of this function and the spectral position this minimum is located at is the first mode that starts to oscillate. If α is a constant, then the minimum is at $\nu = \nu_0$, where $g(\nu_0) = g$ is the maximum of $g(\nu)$, which consequently reproduces Eq. (1.11).

1.5.1 Gain Spectrum

For a QCL with a single radiative transition (between levels 3 and 2) the peak gain coefficient has the form [16]

$$g = \tau_3(1 - \frac{\tau_2}{\tau_{32}})\frac{4\pi e z_{32}^2}{\lambda_0 \varepsilon_0 n_{\text{eff}} L_p 2\gamma_{32}}, \tag{1.14}$$

where τ_3 and τ_2 are the total lifetimes of the upper and lower laser states, τ_{32} is the scattering time between the two, λ_0 is the wavelength in vacuum, ε_0 is the vacuum dielectric constant, n_{eff} is the effective modal index of refraction of the waveguide, z_{32} is the optical dipole matrix element connecting the upper and lower laser states, L_p is the thickness of one period of active region, and γ_{32} is the linewidth of the transition (in units of energy), to be determined theoretically or as the FWHM of the electroluminescence spectrum.

Assuming a homogeneously broadened gain coefficient function with a width γ_{32}, whose peak is located at ν_0 and has the value given in Eq. (1.14), its full form is

$$g(\nu) = g\frac{(\gamma_{32}/2)^2}{(hc)^2(\nu - \nu_0)^2 + (\gamma_{32}/2)^2}, \tag{1.15}$$

where Planck's constant h and the speed of light c were inserted to make the Lorentzian a function of wavenumber.

The (unsaturated) gain is connected to the gain coefficient function and the pump current density J through

$$\gamma_0(\nu) = Jg(\nu), \tag{1.16}$$

which reproduces Eq. (1.12) when $\alpha = \text{const}$, $J = J_{th}$, and $\nu = \nu_0$.

The unsaturated gain, or *small-signal gain*, is the value the gain has in the absence of light in the resonator, since light causes the gain to saturate. This is the gain the amplifier has when the pump rate is smaller than the lasing threshold. But this is also the gain the amplifier has in a brief transient time span during the onset of the laser operation or at the beginning of each laser pulse, i.e. before the resonator fills up with light and saturating the

gain. Thus the effective resonator gain (in the simplest case of the FP oscillations) has the following pump current dependence:

$$\gamma = \begin{cases} \gamma_0 & \text{if } J \leq J_{th} \\ \frac{\alpha}{\Gamma} & \text{if } J > J_{th} \end{cases}$$

If there are multiple transitions, caused either intentionally within one period of the active region, or unintentionally due to drift in the growth parameters and therefore layer thicknesses, then the small signal gain generalizes to

$$\gamma_0(J,\nu) = J\sum_i g_i \frac{(\gamma_{32,i}/2)^2}{(hc)^2(\nu-\nu_{0,i})^2 + (\gamma_{32,i}/2)^2}, \tag{1.17}$$

which can be restated to give the more general form of Eq. (1.12)

$$\gamma_0(J_{th}^{FP},\nu) = \frac{\alpha^{FP}}{\Gamma}\sum_i a_i \frac{(\gamma_{32,i}/2)^2}{(hc)^2(\nu-\nu_{0,i})^2 + (\gamma_{32,i}/2)^2}, \tag{1.18}$$

where $a_i \leq 1$ are weight factors of the different transitions (their magnitude is limited by the fact that the highest peak determines the threshold for FP oscillation).

1.5.2 Losses in a Littrow External Cavity

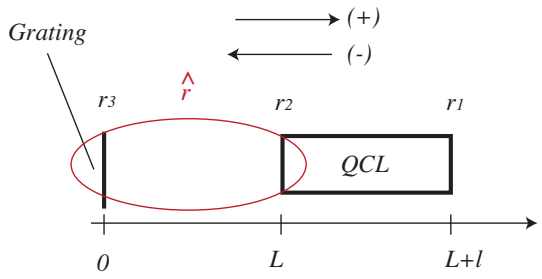


Figure 1.10 *Nomenclature of the symbols used in this section.*

The losses of the external cavity with a grating as the reflector can be calculated using a coupled-cavity model given in [20, 22]. This model combines the partial reflections from the intra-cavity facet r_2 (with non-ideal AR-coating) and the wavelength-dependent reflections from the grating to a complex, phase-dependent, effective amplitude reflection coefficient of the External Cavity

$$\hat{r}(\nu,\nu_g) = r_2^{(-)} - r_G \frac{\eta^2 e^{2iL\beta_{\text{air}}} t_2^{(-)} t_2^{(+)}}{\eta^2 e^{2iL\beta_{\text{air}}} r_2^{(+)} r_G - 1}, \tag{1.19}$$

where $\nu_g(\theta)$ is the central wavelength of the grating-selected reflection band at tuning angle θ, $r_2^{(\pm)}$ and $t_2^{(\pm)}$ are the Fresnel reflection and transmission coefficients of the intra-cavity facet, respectively, in the forward and backward directions (for a sketch of the nomeclature in use, see Fig. 1.10), η is the net (amplitude) coupling efficiency each way due to the intra-cavity optics, $\beta_{\text{air}} = \frac{2\pi}{\lambda_0}$ is the free-space propagation constant, L is the effective free-space

path length between the intra-cavity facet and the grating, and r_G is the grating reflection coefficient.

With this, the overall loss according to Eq. (1.7) is given by

$$\bar{\alpha}(\nu,\nu_g) = \frac{\alpha(\nu,\nu_g)}{\Gamma} = \frac{1}{\Gamma}\left(\alpha_w + \frac{1}{l}\ln\left|\frac{1}{r_1^{(+)}\hat{r}(\nu,\nu_g)}\right|\right), \tag{1.20}$$

where l is the length of the QCL waveguide and $r_1^{(+)}$ is the Fresnel amplitude reflection coefficient of the extra-cavity facet in the forward direction. The phenomenological confinement factor Γ was included here to connect Eq. (1.7) with Eq. (1.11) and $\bar{\alpha}$ is the effective loss that includes the overlap factor and is equal to the threshold gain during laser operation.

The grating reflection coefficient in Eq. (1.19) is

$$r_G = U(\lambda,\theta_B,\tilde{P})\sum_{m=-N/2}^{m=N/2}\tilde{W}(m)e^{i4\pi dm\sin(\theta)/\lambda}, \tag{1.21}$$

where $U(\lambda,\theta_B,P)$ is the effective diffraction function of each grating tooth that depends on the wavelength λ, grating tilt angle θ, grating blaze angle θ_B, and E-field polarization $\tilde{P}$ (TE or TM). $\tilde{W}(m)$ is the normalized amplitude of the electric field at the mth grating tooth, d is the grating period and $N+1$ is the total number of grating teeth illuminated by the beam. The diffraction function can be approximated by

$$U(\lambda,\theta_B,\mathrm{TE}) \approx \sqrt{\eta_G}\frac{\bar{s}^2\sin^2[(\theta-\theta_B)/s]}{(\theta-\theta_B)^2}, \tag{1.22}$$

where $\bar{s}$ is a parameter describing the width of the grating diffraction efficiency near the blaze angle θ_B and η_G is the diffraction efficiency at the blaze angle. If $W(m) = 1/(N+1)$ in Eq. (1.21) is treated as a constant (equivalent to a rectangular beam cross section with a constant beam profile), the sum becomes

$$\sum_{m=-N/2}^{m=N/2}\tilde{W}(m)e^{i4\pi dm\sin(\theta)/\lambda} = e^{-i\pi q}\frac{\sin(\pi(N+1)q)}{(N+1)\sin(\pi q)}, \tag{1.23}$$

where $q = 2d\sin(\theta)/\lambda$. When $\tilde{W}(m)$ is not a constant, as is the general case, since the beam impinging on the grating has a non-constant profile, a more general expression for the sum in Eq. (1.21) is

$$\int_{-\infty}^{\infty}\Xi\left(N\frac{q-\zeta}{2}\right)\frac{\sin(\pi(N+1)\zeta)}{\sin(\pi\zeta)}d\zeta, \tag{1.24}$$

where $\Xi(\kappa)$ is the spatial Fourier transform of the beam profile

$$\Xi(\kappa) = \int_{-\infty}^{\infty}F(x)e^{i2\pi\kappa x}dx, \tag{1.25}$$

where $F(x)$ is the normalized electric field profile, where the coordinate x is scaled with the beam half-width on the grating (the full beam thus extends from $x=-1$ to 1 with a width

of 2).

Using the approximation made for Eq. (1.23) to be valid and assuming the that the blaze angle lies very close to the Littrow angle, so that $U(\lambda, \theta_B, \mathrm{TE}) \approx \sqrt{\eta_G}$, the grating reflection coefficient becomes

$$r_G = \sqrt{\eta_G} e^{-i\pi q} \frac{\sin(\pi(N+1)q)}{(N+1)\sin(\pi q)}. \tag{1.26}$$

The grating (intensity) reflectivity $|r_G|^2$ is plotted in Fig. 1.11 for a grating with a groove distance of 1/300 mm and a grating reflectivity $\eta_G = 85\%$, tuned to a Littrow angle of $\theta = 41.36°$, and a beam diameter of 5 mm. The response from the grating has the shape of the first-order maximum of a multi-slit diffraction pattern with a peak reflectivity of 85%, strongly suppressed satellite peaks, and a width that is proportional to the number of illuminated grooves.

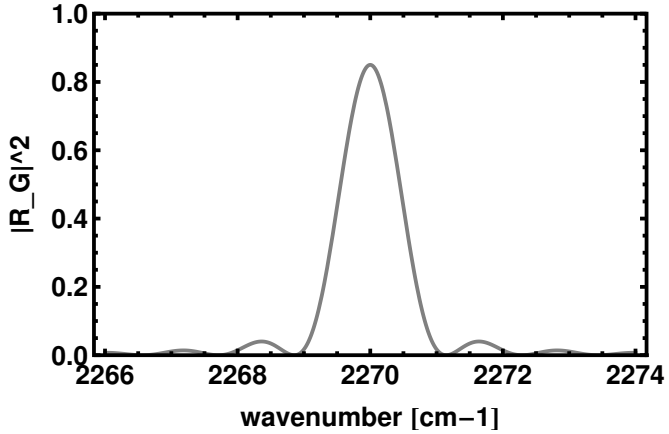


Figure 1.11 *Grating reflectivity for a Littrow angle of* $\theta = 41.36°$.

Using Eq. (1.26) as the reflectivity of the grating, a round-trip (intensity) coupling efficiency of $\eta^4 = 25\%$, and a cavity length of $L = 23$ mm, the effective reflectivity of the external cavity, $|\hat{r}|^2$ of Eq. (1.19), can be calculated. This is plotted in Fig. 1.19 for three different residual reflectivities of the AR-coated intra-cavity facet. The gray line is for the case of an uncoated facet with an as-cleaved Fresnel reflectivity of $R_2 = |r_2^{(-)}|^2 = 28\%$, the blue line is for a reflectivity of 1.5%, and the red line is for 0.1%. Note, that far away from the Littrow wavenumber, the asymptotical reflectivity of the external cavity is just the reflectivity of the intra-cavity facet. The oscillations are due to the coupled-cavity effect, i.e. whether the partial wave reflected from the facet and the partial wave reflected from the grating are in phase or out of phase, depending on whether the free space path length between the facet and the grating is an integral multiple of half the wavelength. The oscillations consequently become smaller with better AR-coatings, since the role of the partial reflection from the intra-cavity facet diminishes with lower residual reflectivities.

The round-trip loss within the external cavity (including the QCL) according to Eq. (1.20) for a QCL with $\alpha_w = 4\,\mathrm{cm}^{-1}$, $\Gamma = 1$, and a length $l = 6$ mm is plotted in Fig. 1.13 for

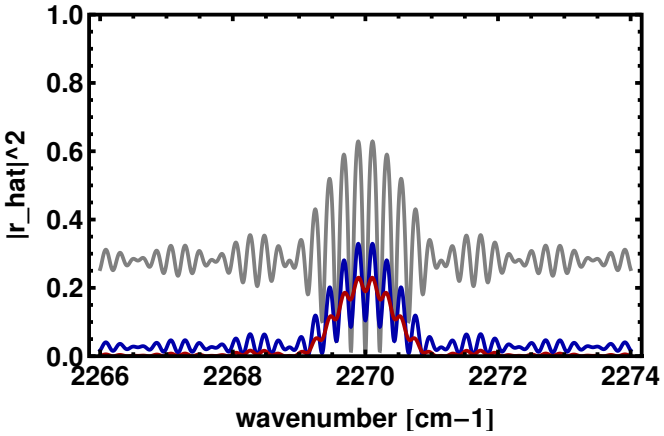


Figure 1.12 *Reflectivity of the external cavity* $|\hat{r}(\nu, \nu_g = 2270\,\mathrm{cm}^{-1})|^2$ *at three different values of the residual reflectivity of facet 2.*

residual reflectivity of the coated facet of 1.5%. The loss has to be outweighed by the gain for laser action to occur.

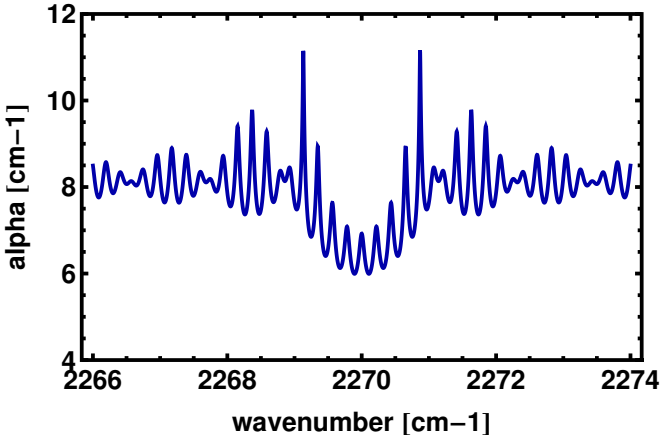


Figure 1.13 *Wavenumber-dependent loss of the external cavity with a grating tuned to* $\nu_g = 2270\,\mathrm{cm}^{-1}$.

Figure 1.14 shows the losses plotted in Fig. 1.13 on a coarse scale. Ignoring the fine detail of the loss at the Littrow wavenumber for now, the loss has the following structure. At the grating-selected wavenumber ν_g, the cavity has a reduced loss value, here $\alpha^{EC}/\Gamma = 6.2\,\mathrm{cm}^{-1}$ (with $\Gamma = 1$). At wavenumbers away from ν_g, the losses have a higher value given by the residual reflectivity of the AR coating, $\alpha^{FP}/\Gamma = 8.15\,\mathrm{cm}^{-1}$ (with $\Gamma = 1$), where the FP stands for Fabry-Perot.

If the AR coating is very good and we assume the effective grating reflectivity to be independent of tuning angle (which is equivalent to saying the tuning angle is near the blaze angle), then the feedback maximum (exactly at the Littrow angle) is also independent of Littrow angle. With this, α^{EC} is also a constant with respect to tuning angle. But α^{FP} is also a constant, as it is the asymptotical value the losses reach far away from the Littrow angle. So in a coarse view (disregarding the modal fine structure), the external cavity losses

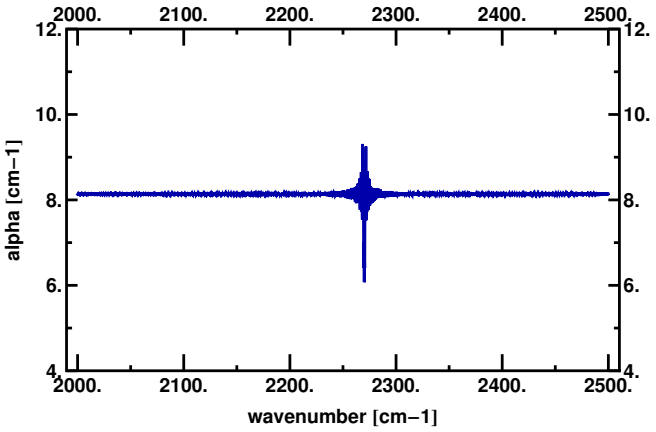


Figure 1.14 *Wavenumber-dependent loss of the external cavity with a grating tuned to* $\nu_g = 2270\,\mathrm{cm}^{-1}$ *on a wider scale.*

can be treated as a discontinuous distribution that has the values

$$\alpha(\nu, \nu_g) = \begin{cases} \alpha^{EC} & \text{if } \nu = \nu_g \\ \alpha^{FP} & \text{if } \nu \neq \nu_g \end{cases}$$

From Eqs. (1.20) and (1.12) it is clear that these approximations result in

$$\alpha^{FP} = \alpha_w + \frac{1}{l} \ln \left| \frac{1}{r_1^{(+)}} \right| + \frac{1}{l} \ln \left| \frac{1}{r_2^{(-)}} \right| \tag{1.27}$$

and

$$\alpha^{EC} = \alpha_w + \frac{1}{l} \ln \left| \frac{1}{r_1^{(+)}} \right| + \frac{1}{l} \ln \left| \frac{1}{r_G} \right|. \tag{1.28}$$

1.5.3 Threshold Condition and Tuning

Figure 1.15 shows plots of gain spectra and losses for different grating angles and pump currents. The gain spectra are calculated using a modified version of Eq. (1.16),

$$\gamma_0(\nu) = g(\nu) \frac{I}{lb}, \tag{1.29}$$

where $I = Jlb$ is the pump current and lb is the cross sectional area normal to the current flow, where l is the length of the QCL active region and b is its width. The red curve is calculated from Eq. (1.29) using a value of $\frac{g}{lb} = 1.36/(\mathrm{A\ cm})$, a current of $I = 4.4$ A, a central wavenumber of $\nu_0 = 2270\,\mathrm{cm}^{-1}$, and a width $\gamma_{32}/(hc) = 297\,\mathrm{cm}^{-1}$ (h is Planck's constant), the orange curve is at $I = 6.0$ A. The blue curve is the same as is plotted in Fig. 1.13 (with a residual reflectivity of the AR coating of 1.5% and a grating tuned to $\nu_g = 2270\,\mathrm{cm}^{-1}$). The waveguide loss $\alpha_w = 4\,\mathrm{cm}^{-1}$, the QCL length is $l = 6$ mm and the other facet is uncoated. The gray curve uses the same parameters, except the grating is tuned to $\nu_g = 2180\,\mathrm{cm}^{-1}$. The black line represents the Fabry-Perot losses of the same laser in its uncoated state without feedback from the external grating.

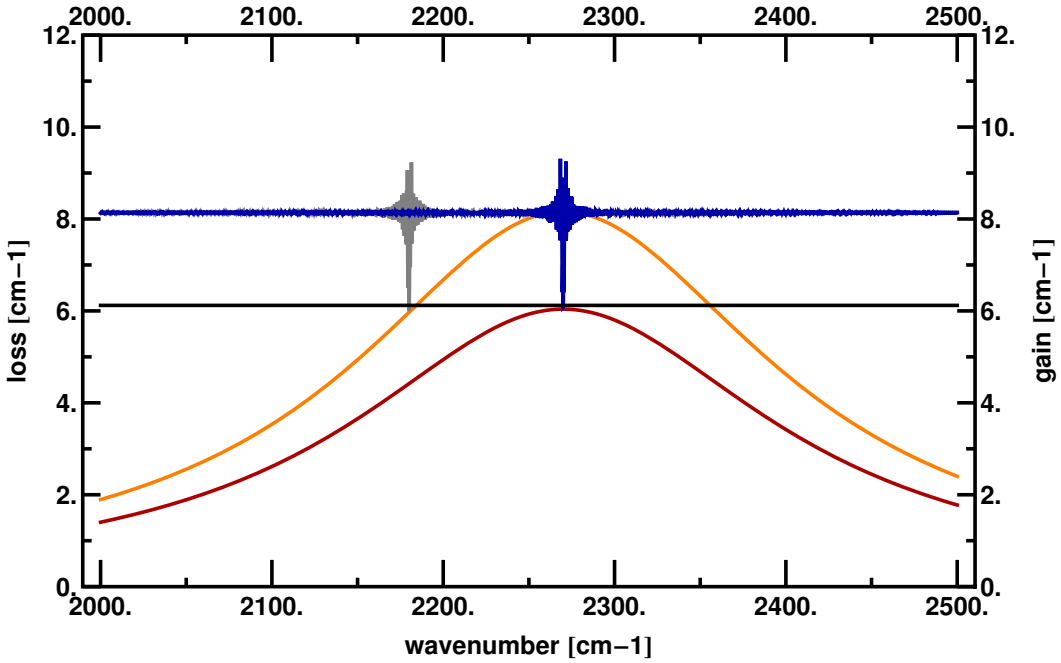


Figure 1.15 *Plots of gain spectra and losses for different grating angles and pump currents. Red: Gain spectrum for a pump current of $I = 4.4$ A. Orange: Gain spectrum for a pump current of $I = 6.0$ A. Blue: EC loss curve for 1.5% AR coating, grating at $\nu_g = 2270\,\mathrm{cm}^{-1}$). Gray: EC loss curve for 1.5% AR coating, grating at $\nu_g = 2180\,\mathrm{cm}^{-1}$). Black, Fabry-Perot loss of uncoated QCL.*

The Fabry-Perot losses (black) add up to a value of $6.2\,\mathrm{cm}^{-1}$, thus, pumping the uncoated QCL at $I = 4.4$ A (red) makes the top of the gain curve *touch* the loss value. This is the condition for the threshold current, which is therefore $I_{th}^{FP}(\text{uncoated}) = 4.4$ A and the laser starts to oscillate at $2270\,\mathrm{cm}^{-1}$. Coating the QCL facet raises the losses without external feedback to $8.15\,\mathrm{cm}^{-1}$ as can be seen from the asymptotical value of the blue curve far from the grating dip. Pumping the QCL with $I = 6.0$ A (orange) makes the coated QCL's gain reach this loss value, and consequently the threshold current for Fabry-Perot oscillation of the coated QCL is $I_{th}^{FP}(\text{coated}) = 6.0$ A. Pumping the coated QCL below this value ensures oscillation on the grating-selected mode only, and the maximum pump rate for pure EC oscillation is just under 6.0 A.

The dip of the EC losses is also at $6.2\,\mathrm{cm}^{-1}$. Thus the effective feedback of the external cavity is very similar to the feedback of the uncoated QCL facet $|\hat{r}(\nu = \nu_g)|^2 \simeq 28\%$. Thus, if the grating is tuned to the maximum of the gain, the laser also starts to oscillate at $I_{th}^{EC,2270}(\text{coated}) = 4.4$ A. However, at this pump rate the tuning range is zero, since if the grating is slightly detuned from the gain max, there is no point where the loss touches the gain, thus the laser seizes operation. Pumping the laser at the maximum allowable current for pure EC operation (orange), however, raises a substantial part of the gain spectrum above $6.2\,\mathrm{cm}^{-1}$. Now even the loss curve of the grating tuned to $2180\,\mathrm{cm}^{-1}$ (gray) touches the gain. If it is tuned any further, the laser shuts off. Thus the tuning range is $\Delta\nu = 2 \times (2270 - 2180)\,\mathrm{cm}^{-1} = 180\,\mathrm{cm}^{-1}$ and the threshold current for the grating tuned to $2180\,\mathrm{cm}^{-1}$ is

$I_{th}^{EC,2180}$(coated) = 6.0 A. As will be shown in Section 1.5.5, the output power is proportional to the difference between gain and loss, thus at the center of the tuning range the output power is also at a maximum. Due to gain clamping (discussed in more detail in Section 2.5.4), the red curve also represents the saturated gain of the laser pumped at 6.0 A with a grating tuned to 2270 cm^{-1}.

To illustrate the impact of the quality of the AR coating on the tuning range and the maximum output power, Figure 1.16 shows a comparison of the situations with two different residual reflectivities. The gray and the orange curves are the same as in Fig. 1.15, i.e. the EC loss for an AR coating of with 1.5% residual reflectivity, and its FP-limited gain, respectively. The black and the purple line show the same for a residual reflectivity of 0.1%. The limiting pump current is increased from 6.0 A to 7.95 A. The loss curves have the Littrow grating tuned to their respective tuning range limits, which is2180 cm^{-1} for the first AR coating and 2144 cm^{-1} for the latter. Thus the tuning range is increased by 40%. Tuning the grating to the center of the gain resulted in a threshold current of 4.4 A, as mentioned before. Assuming a linear P-I characteristic, a pump limit of 7.95 A results in a maximum output power that is increased by a factor of 2.25 compared with a pump limit of 6.0 A.

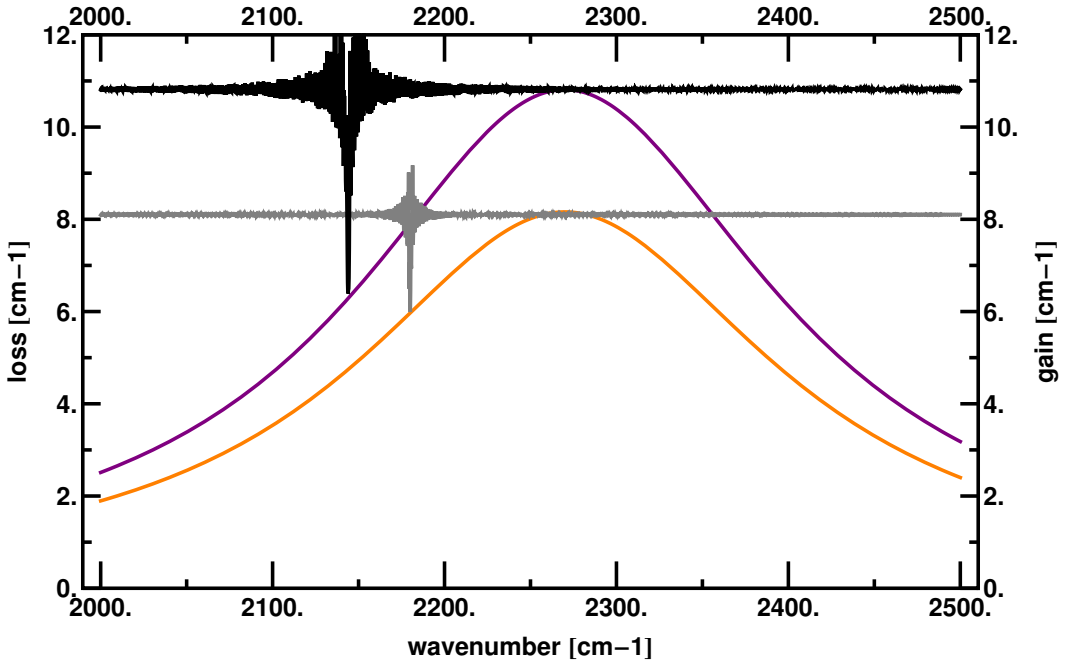


Figure 1.16 *Plots of gain and losses for different AR coatings and pump currents. Orange: Gain spectrum for a pump current of $I = 6.0$ A. Purple: Gain spectrum for a pump current of $I = 7.95$ A. Gray: EC loss curve for 1.5% AR coating, grating at $\nu_g = 2180\,\mathrm{cm}^{-1}$). Black: EC loss curve for 0.1% AR coating, grating at $\nu_g = 2144\,\mathrm{cm}^{-1}$).*

1.5.4 Tuning Range

Although the tuning range has partially been discussed in the previous section, this section gives a more formal approach.

In case there is one homogeneously broadened laser transition, the gain is a Lorentzian with width γ_{32} and a peak value that is proportional to the pump current density J. From Fig. 1.15 it is clear that to assure oscillation exclusively on the EC modes, the QCL has to be pumped below the FP threshold. But the FP threshold is just reached if the top of the gain curve meets α_{FP} (for $\Gamma = 1$, otherwise α_{FP}/Γ). Thus, from Eq. (1.13),

$$J_{th}^{FP} = \frac{\alpha^{FP}}{\Gamma g(\nu_0)}. \tag{1.30}$$

The Fabry-Perot threshold is a constant of tuning angle and the associated Littrow wavenumber ν_g, since FP oscillation always starts at the top of the gain spectrum ν_0. The threshold for EC oscillation, however, is obviously a function of ν_g and is given by

$$J_{th}^{EC}(\nu_g) = \frac{\alpha^{EC}}{\Gamma g(\nu_g)}. \tag{1.31}$$

To reach maximum power and tuning range, it is best to pump the laser a small amount below the FP threshold, at $J_{th}^{FP(-)}$ (however, in the following the index $(-)$ is dropped for readability). In this case, according to Eq. (1.16) and Eq. (1.15), the small-signal gain is just

$$\gamma_0(J_{th}^{FP}, \nu) = J_{th}^{FP} g \frac{(\gamma_{32}/2)^2}{(hc)^2(\nu - \nu_0)^2 + (\gamma_{32}/2)^2}, \tag{1.32}$$

which is just

$$\gamma_0(J_{th}^{FP}, \nu) = \frac{\alpha^{FP}}{\Gamma} \frac{(\gamma_{32}/2)^2}{(hc)^2(\nu - \nu_0)^2 + (\gamma_{32}/2)^2}. \tag{1.33}$$

But now the tuning range is just the range of Littrow wavenumbers, for which the EC losses are smaller than the small-signal gain (the saturated gain will clamp to the EC loss value at the Littrow wavenumber)

$$\frac{\alpha(\nu = \nu_g)}{\Gamma} \equiv \frac{\alpha^{EC}}{\Gamma} < \gamma_0(J_{th}^{FP}, \nu_g). \tag{1.34}$$

The limits of the tuning range are therefore given by

$$\nu_g^{1,2} = \nu_0 \pm \frac{\gamma_{32}}{2hc} \sqrt{\frac{\alpha^{FP}}{\alpha^{EC}} - 1}. \tag{1.35}$$

Finally, the tuning range is given by

$$\Delta\nu_g = \frac{\gamma_{32}}{hc} \sqrt{\frac{\alpha^{FP}}{\alpha^{EC}} - 1}. \tag{1.36}$$

The two quantities determining the tuning range are therefore γ_{32}, as an intrinsic property of the QCL, and the mixed intrinsic-extrinsic ratio $\frac{\alpha^{FP}}{\alpha^{EC}}$, which is increased by increasing

α^{FP} through better anti-reflection coatings, and by increasing the EC feedback by using high-reflectivity gratings close to the blaze angle and reducing the coupling losses by using accurately positioned, AR-coated collimating lenses with high numerical apertures. Also, the ratio is lowered by lowering the intrinsic waveguide losses α_w and by lowering the reflectivity of the other facet.

Using Eq. (1.30) and Eq. (1.31) with the grating tuned to the gain maximum, $\nu_g = \nu_0$, Eq. (1.36) can be restated as

$$\Delta\nu_g = \frac{\gamma_{32}}{hc}\sqrt{\frac{J_{th}^{FP}}{J_{th}^{EC}(\nu_g = \nu_0)} - 1}. \tag{1.37}$$

Another interesting aspect to consider is the temperature-dependence of the tuning range of Eq. (1.36).

Since α_w in α^{FP} and α^{EC} is virtually temperature independent, as are the various reflectivities, the square root in Eq. (1.36) is temperature-independent, and the only temperature dependence lies within $\gamma_{32}(T)$. This is quite remarkable, since the threshold currents in Eq. (1.36) are highly temperature-dependent. In fact, there is an exponential dependence normally stated using the phenomenological characteristic temperature T_0 [16] through

$$J_{th}(T) \approx J_{th}(0)e^{\frac{T}{T_0}}. \tag{1.38}$$

However, the intrinsic temperature parameter T_0 does not change with extrinsic measures such as coating the facet with an AR coating or running the laser in an EC. Therefore, the tuning range becomes

$$\Delta\nu_g(T) \approx \frac{\gamma_{32}(T)}{hc}\sqrt{\frac{J_{th}^{FP}(T=0)e^{\frac{T}{T_0}}}{J_{th}^{EC}(\nu_g = \nu_0, T=0)e^{\frac{T}{T_0}}} - 1}, \tag{1.39}$$

so the explicit temperature dependence cancels out:

$$\Delta\nu_g(T) \approx \frac{\gamma_{32}(T)}{hc}\sqrt{\frac{J_{th}^{FP}(T=0)}{J_{th}^{EC}(\nu_g = \nu_0, T=0)} - 1}. \tag{1.40}$$

So the only temperature dependence is through $\gamma_{32}(T)$. But $\gamma_{32}(T)$ grows with increasing temperatures, which means that the tuning range is larger for higher temperatures. The effect is however reduced in case there are multiple transitions with overlapping gains and other effects.

1.5.5 Power Spectrum and Maximum Power of the Tuned ECQCL

Gain saturation in a laser (this will be discussed in more detail in Section 2.5.4) is due to the internal light causing stimulated emission, thereby reducing the population of the upper level and with it the overall gain while increasing the photon number. The balance between internal photon flux and gain is given by the expression

$$\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \phi(\nu)/\phi_s(\nu)}, \tag{1.41}$$

where $\gamma(\nu)$ is the saturated gain, $\phi(\nu)$ is the internal photon flux density, $\phi_s(\nu)$ is the saturation photon flux density, and $\gamma_0(\nu)$ is the small-signal gain, i.e. the gain at the very beginning of the pumping process, before the gain reaches an equilibrium with the photon flux density. Rewriting this in terms of the photon flux density yields

$$\phi(\nu) = \phi_s(\nu)\left(\frac{\gamma_0(\nu)}{\gamma(\nu)} - 1\right). \tag{1.42}$$

Since the saturated gain at the Littrow wavenumber $\gamma(\nu = \nu_g)$ is clamped at the EC value α^{EC}/Γ, and is therefore a constant, Eq. (1.42) becomes

$$\phi(\nu_g) = \frac{\phi_s(\nu_g)\Gamma}{\alpha^{EC}}\left(\gamma_0(\nu_g) - \frac{\alpha^{EC}}{\Gamma}\right), \tag{1.43}$$

which makes clear that the power output of the EC is proportional to the difference between small-signal gain at the Littrow wavenumber and the EC loss. Therefore the EC can be used as a tool to analyze the gain spectrum without having to resort to electroluminescence measurements. This becomes even clearer in the following.

If the QCL is pumped just below threshold for FP oscillation, the small-signal gain is given by Eq. (1.33). Consequently, provided ν_g is within the tuning range, Eq. (1.43) at $\nu = \nu_g$ reads

$$\phi(\nu = \nu_g) = \phi_s(\nu_g)\left(\frac{\alpha^{FP}}{\alpha^{EC}}\frac{(\gamma_{32}/2)^2}{(hc)^2(\nu_g - \nu_0)^2 + (\gamma_{32}/2)^2} - 1\right). \tag{1.44}$$

Eq. (1.44) provides the shape of the power spectrum as the grating is tuned, since the output power (out of facet i) is proportional to the internal photon flux

$$P_i = \frac{\phi}{2}(1 - R_i)A\hbar\omega, \tag{1.45}$$

where A is the facet area and R_i the reflectivity of facet i, and $\hbar\omega$ is the energy of a photon. This relation is strictly only valid in case the reflectivities are approximately 1, as discussed in detail in Section 2.5.4. As a sanity check, solving Eq. (1.44) for the the values of ν_g that make ϕ vanish, results in the limits of the tuning range Eq. (1.35).

Of course, Eq. (1.44) is only valid, if there is only one laser transition, in this case the gain spectrum of the tuned ECQCL will have the shape of a truncated Lorentzian. However, often there is more structure to the tuned power spectrum due to there being more than one

transition energy either due to multiple transitions or drift of the growth parameters. In that case, the small-signal gain in Eq. (1.43) has the shape given in Eq. (1.33) which translates linearly into Eq. (1.44).

Tuning the grating to the center of the gain region yields the maximum photon flux

$$\phi(\nu_g = \nu_0) = \phi_s(\nu_g) \left(\frac{\alpha^{FP}}{\alpha^{EC}} - 1 \right), \tag{1.46}$$

which states that the maximum power grows linearly with the ratio $\frac{\alpha^{FP}}{\alpha^{EC}}$.

An interesting consequence of Eq. (1.46) is that since the ratio $\frac{\alpha^{FP}}{\alpha^{EC}}$ is temperature-independent, the maximum achievable power with an ECQCL in pure EC operation is independent of temperature. Since the equilibrium temperature of the active region is also affected by the pump current due to Joule heating, the duty cycle of a pulsed QCL also affects the average temperature of the active region during the pulse. But since the output power is independent of temperature, the maximum output in in pure EC operation during a pulse is also independent of duty cycle. This will be experimentally investigated in Section 4.4.

1.5.6 Modal Fine Structure and Mode-Hopping

The gain and the loss can both be plotted over the continuum of wavenumbers (Figs. 1.13 and 1.15). However, in a laser, according to Eq. (1.6), only a discrete set of wavenumbers fulfill the resonance condition required for the positive feedback loop that enables laser action. These are the *modes* of the resonator. For a Fabry-Perot laser, the reflection coefficients r_1 and r_2 are normally constant over the wavenumber range of interest. In case of the Fresnel reflectivities of (nearly) transparent materials, they are also (nearly) real, thus the argument is either 0 or $\pm\pi$, and the modes are equidistant and can be numbered such as in Eq. (1.9). For an EC, Eq. (1.6) translates into

$$2l\beta(\nu) + \arg(r_1) + \arg(\hat{r}(\nu, \nu_g)) = 2m\pi, \tag{1.47}$$

where $\arg(\hat{r}(\nu, \nu_g))$ is now a highly oscillatory function (within the grating's reflective band) because of the large distance traveled to the grating and back, which enters the argument through the terms form $e^{2iL\beta_{\text{air}}} = e^{4i\pi\frac{L}{\lambda}}$.

These modes can be marked in the loss curve of the EC laser, e.g. Fig. 1.13, to give Fig. 1.17. The Fabry-Perot modes – i.e. the modes far from the grating wavenumber, where the grating reflectivity is negligible and thus the free-space contribution is as well – are spaced at $1/(2nl)$, where l is the chip length. The EC modes, i.e. the modes close to the grating wavenumber, are spaced at $1/(2(nl + L))$ since the resonator length is now the chip length plus the free-space path. The loss curve, however, oscillates with a period $1/(2L)$, since its oscillatory part is only determined by the distance between facet and grating. Therefore, the loss curve is intersected by the allowed modes in a complicated manner, that depends on the exact values ν_g, l, and L, with no straight-forward rule about when it is the dips or the peaks

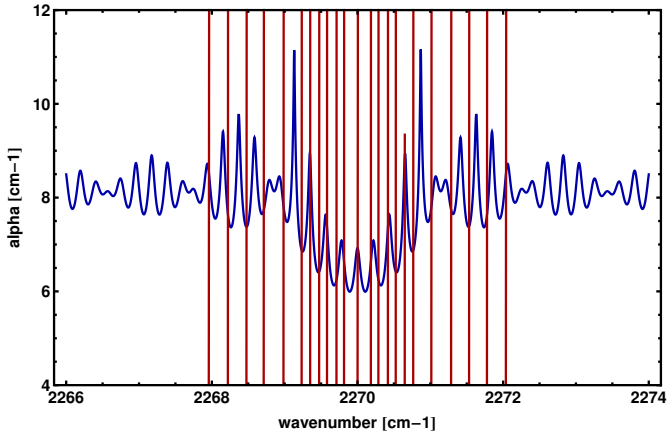


Figure 1.17 *Wavenumber-dependent loss of the external cavity with an AR coating reflectivity of* 1.5%*, with a grating tuned to* $\nu_g = 2270\,\mathrm{cm}^{-1}$*). The allowed modes (only plotted between* $2268\,\mathrm{cm}^{-1}$ *and* $2272\,\mathrm{cm}^{-1}$*) are marked with vertical red lines.*

that are intersected. This rule can be given for the limiting case that the free-space length is much larger than the chip length. In this case, when the grating is tuned to coincide with an FP mode, the EC modes coincide with dips of the loss curve. When the grating is tuned half way between FP modes, the the EC modes coincide with peaks of the loss curve.

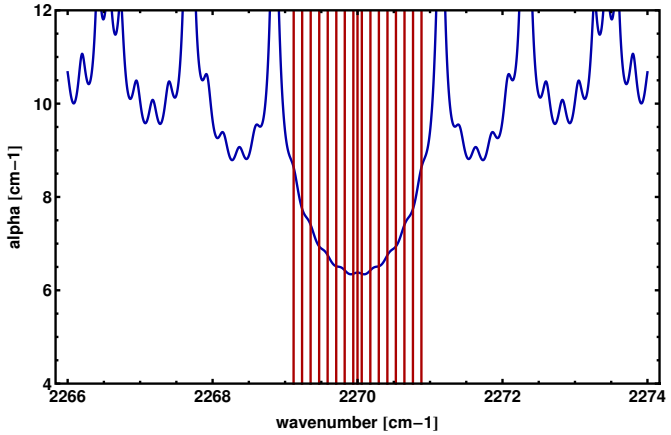


Figure 1.18 *Wavenumber-dependent loss of the external cavity with an AR coating reflectivity of* 10^{-4}*, with a grating tuned to* $\nu_g = 2270\,\mathrm{cm}^{-1}$*). The allowed modes (only plotted between* $2269\,\mathrm{cm}^{-1}$ *and* $2271\,\mathrm{cm}^{-1}$*) are marked with vertical red lines.*

Since the loss curve is identical to the threshold gain curve, its value at a particular mode is the threshold for that mode. The mode with the lowest threshold will be the dominant mode for laser action, and the others can be disregarded after a very brief transient phase after the onset of the pump current. For instance, it can be seen in Fig. 1.17 that the mode at exactly $2270.0\,\mathrm{cm}^{-1}$ is strongly suppressed, while the third mode toward the left and the right (counting from the central mode), at $2269.7\,\mathrm{cm}^{-1}$ and $2270.3\,\mathrm{cm}^{-1}$, are the ones that start to oscillate first. The higher the lowest intersection lies for a given grating angle, the lower the overall threshold of the laser and the stronger its emission at a given pump current. Thus, as the grating is tuned and the EC modes intersect the loss curve at varying

loss values, the overall intensity oscillates. The phenomenon that the selected mode does not tune continuously, but "hops" to wherever the threshold is lowest, is called *mode-hopping*. As is clear from Fig. 1.12, the oscillations of the effective EC reflectivity are damped with better AR coatings, since the coupled-cavity effect is suppressed. The extreme case for an AR coating with a residual reflectivity of 10^{-4} is plotted in Fig. 1.18. Since the loss curve is much smoother now, it is clear that if the grating is tuned, the lowest threshold stays close to the minimum which is close to the grating-selected wavenumber.

To avoid mode hopping different approaches can be taken. One uses a long external cavity in conjunction with a long chip driven in pulsed mode. Here, the oscillatory heating of the active region chirps the wavelength across the densely-spaced EC modes through variation of the chip's refractive index. The drawback of this approach is that the emission is multimode and the linewidth is relatively large, since it is the envelop of the modes. However, the great advantage is simplicity, and the overall linewidth can be limited by using a very large beam diameter to make the grating-envelope narrow. This approach is taken in Chapter 3.4.

The second approach combines a good AR coating with a QCL driven in continuous-wave (CW) mode and a grating with a precisely positioned pivot point for the grating. This approach is given in [127]. The pivot point is chosen so that as the grating is rotated, the cavity length varies precisely in such a way as to leave the *same* mode in the minimum of the grating-selected band. Thus, the angle of the grating and the overall length of the cavity have to tune in synchron, so that a single grating-selected mode "surfs" along on the loss-minimum as it is tuned. The advantage of this approach is that it allows mode-hop-free broadband tuning with a chirp-less CW linewidth. The drawback is that the setup has to be immensely stable, since if the pivot point moves by as much as a micron, mode-hops reappear during tuning.

Two other approaches to avoid mode-hopping during CW operation take the very obvious route of adjusting the cavity length at each tuning step by maximizing the output with a closed feedback loop. One adjusts the effective chip length by thermally altering the refractive index. For this, the heat sink temperature of the QCL needs to be adjusted at each tuning step, which is very slow and not practical for actual continuous tuning, only for stepwise continuous tuning. The other adjusts the grating's pivot point by mounting the entire grating on a piezo actuator. Although a piezo crystal can actuated very quickly, if the grating is rocked too abruptly, it starts to vibrate at its mechanical resonance. Since a grating is a very bulky load, the resonance frequency is very low and in the worst case, its mechanical resonance is close to that of the entire setup, which can lead to serious damage. Thus this approach is also neither fast, nor stable. Quite oppositely, it is very technology intensive, and therefore even less fail-safe. Despite all of this, it is quite commonly used in commercial External Cavity QCLs as discussed in Section 1.3.

2

The Anti-Reflection coating of the QCL facet

2.1 Introduction

The facet of the QCL facing the external reflector of the External Cavity – this is normally called the *intra-cavity* facet – needs to receive an optical coating that eliminates its inherent Fresnel reflectivity. Without an anti-reflection (AR) coating, the two distinct spatial regions of the External Cavity resonator, the QCL chip and the free-space region, behave as two separate cavities that are coupled through the intra-cavity facet's finite reflectivity, see Section 1.5.2. This coupling results in unwanted interaction that leads to strongly fluctuating intensities of the output beam as the laser is tuned, or to the extreme case of discrete tuning with gaps between the allowed modes. This is called *mode hopping* and is discussed in detail in Section 1.5.6.

We fabricate these coatings in our labs using reactive magnetron sputtering with quasi in-situ measurement of laser output to determine the quality of the coating during the deposition process. Due to the key role the AR coating plays for EC operation, this thesis dedicates this chapter to its design and fabrication as well as its non-trivial characterization.

2.2 Fundamentals of Optical Coatings: The Matrix Method

The anti-reflection coating is a single layer or a structure of several layers of transparent optical materials with thicknesses on the order of a fraction of the wavelength of the light passed through it. The plain interfaces between these materials – in the simplest single-layer form, the interfaces between the substrate, the layer, and the ambient air – are smooth and abrupt and light passing through them exhibits partial Fresnel reflections. These partial waves interfere with each other in both the forward and backward directions. An AR coating thus is a simple interference filter with the effect that light, preferably of all wavelengths under consideration, interferes constructively in the forward direction and completely destructively

in the reverse. If this is the case, all of the light passes through the structure and none is reflected.

In this section, a simple matrix method to calculate interesting properties such as transmission and reflection through and from a stack of an arbitrary number of thin layers on a substrate will be given. Although derivations are given many times in the literature, I am including one developed for this thesis here, as it combines the original work by Walter Weinstein from 1947 [128] for its clarity while rewriting the solution in terms of the Fresnel coefficients (similar to the standard treatment of the topic by Florin Abelès from 1950 [129, 130], which is outlined by Oliver Heavens in his book of 1960 [131]). Also, much of this thesis – optimization of AR coatings and calculation of reflectivities and transmissivities for the bandpass interference filters – bases on a computer programs written on the basis of this derivation and included in the Appendix C.2 and C.3.

Consider a system of $p+1$ media that is made up of $p-1$ layers on a substrate in an ambient medium (e.g. air). Light is taken to be incident from medium p and not as might be expected from medium 0, because as will be seen later, the solution to the problem can be given in matrix notation where each matrix represents a layer. Thus, adding a layer to the system will have the effect of simply multiplying a matrix with the next index p to the solution.

Refer to Fig. 2.1 for an illustration of the coordinate system and indices used. The paper plane is the plane of incidence and is the x-y-plane where the x-direction is the layer normal and indicating the positive direction of propagation. The z-axis faces toward the reader.

2.2.1 Case I: Light Polarized Parallel to the Plane of Incidence

n_k = refractive index of medium k $(k = 0...p)$
d_k = thickness of medium k $(k = 1...p-1)$
E_p^+ = amplitude of electric field vector in incident beam
E_p^- = amplitude of electric field vector in reflected beam
E_0^+ = amplitude of electric field vector in transmitted beam
E_k^+ = amplitude of electric field vector in beam traveling in positive direction in medium k $(k = 1...p-1)$
E_k^- = amplitude of electric field vector in beam traveling in negative direction in medium k $(k = 1...p-1)$
ϕ_k = angle of incidence in the positive direction in medium k $(k = 1...p-1)$
Y_k^+ = y component of E_k^+ $(k = 0...p)$
Y_k^- = y component of E_k^- $(k = 0...p)$
γ_k^+ = z component of magnetic field vector corresponding to E_k^+
γ_k^- = z component of magnetic field vector corresponding to E_k^-

Plane waves polarized parallel to the plane of incidence in medium k have a wave vector

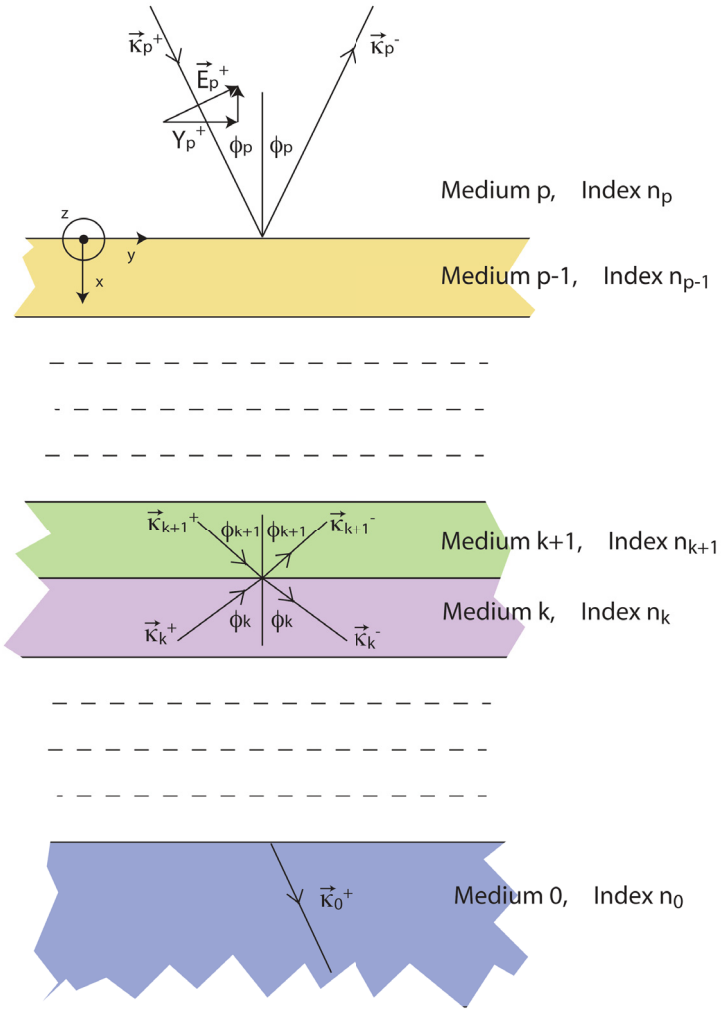


Figure 2.1 *Illustration of the coordinate system and notation used. Indicated for clarity is one electric field vector that is polarized parallel to the plane of incidence (Case I). The magnetic field vector is not drawn.*

(don't confuse the k's):

$$\vec{k}_k^{\pm} = \frac{2\pi n_k}{\lambda}\vec{\kappa}_k^{\pm}$$

where κ is a normalized vector denoting the direction of propagation:

$$\vec{\kappa}_k^{\pm} = \begin{pmatrix} \pm\cos\phi_k \\ \sin\phi_k \\ 0 \end{pmatrix}$$

The electric field vectors

$$\vec{E}_k^{\pm} = \begin{pmatrix} X_k^{\pm} \\ Y_k^{\pm} \\ Z_k^{\pm} \end{pmatrix}$$

thus point in the direction

$$\vec{e}_k^{\pm} = \begin{pmatrix} \mp \sin\phi_k \\ \cos\phi_k \\ 0 \end{pmatrix}.$$

We are only interested in the y-components of the electric field. These read for the two directions ($k = 0...p$):

$$Y_k^+ = \cos\phi_k E_k^+ \exp\frac{2\pi i}{\lambda}(n_k(x\cos\phi_k + y\sin\phi_k) - ct)$$

and

$$Y_k^- = \cos\phi_k E_k^- \exp\frac{2\pi i}{\lambda}(n_k(-x\cos\phi_k + y\sin\phi_k) - ct)$$

$E_k^{\pm}$ can still contain a constant phase factor. Using Maxwell's equations in the CGS-system of units, we know that the electric and magnetic field vectors fulfill the relation

$$\vec{\kappa}^{\pm} \times \vec{H}_k^{\pm} = -n_k \vec{E}_k^{\pm}, \tag{2.1}$$

therefore, the z-components of the magnetic field are

$$\gamma_k^+ = n_k E_k^+ \exp\frac{2\pi i}{\lambda}(n_k(x\cos\phi_k + y\sin\phi_k) - ct)$$

and

$$\gamma_k^- = -n_k E_k^- \exp\frac{2\pi i}{\lambda}(n_k(-x\cos\phi_k + y\sin\phi_k) - ct).$$

The relevant boundary conditions are that the tangential components of the electric and magnetic fields are to be continuous at the p interfaces. This gives

$$\begin{aligned}
Y_0^+ &= Y_1^+ + Y_1^- \quad at \quad x = d_{p-1} + d_{p-2} + ... + d_1 \\
Y_1^+ + Y_1^- &= Y_2^+ + Y_2^- \quad at \quad x = d_{p-1} + d_{p-2} + ... + d_2 \\
&\vdots \\
Y_{p-1}^+ + Y_{p-1}^- &= Y_p^+ + Y_p^- \quad at \quad x = 0 \\
\gamma_0^+ &= \gamma_1^+ + \gamma_1^- \quad at \quad x = d_{p-1} + d_{p-2} + ... + d_1 \\
\gamma_1^+ + \gamma_1^- &= \gamma_2^+ + \gamma_2^- \quad at \quad x = d_{p-1} + d_{p-2} + ... + d_2 \\
&\vdots \\
\gamma_{p-1}^+ + \gamma_{p-1}^- &= \gamma_p^+ + \gamma_p^- \quad at \quad x = 0
\end{aligned} \tag{2.2}$$

Writing these up and canceling factors of the form $\exp(2\pi i/\lambda)(n_k y \sin\phi_k - ct)$ (using Snell's law $n_k \sin\phi_k = n_{k+1}\sin\phi_{k+1}$) and substituting the following symbols

$$
\begin{aligned}
v_k &= (2\pi/\lambda) n_k \cos\phi_k \ldots (k = 0...p) \\
x_k^+ &= \cos\phi_k E_k^+ \exp\{iv_k(d_{p-1} + d_{p-2} + ... + d_k)\} \ldots (k = 1...p-1) \\
x_k^- &= \cos\phi_k E_k^- \exp\{-iv_k(d_{p-1} + d_{p-2} + ... + d_k)\} \ldots (k = 1...p-1) \\
x_0^+ &= \cos\phi_0 E_0^+ \exp\{iv_0(d_{p-1} + d_{p-2} + ... + d_1)\} \\
x_p^+ &= \cos\phi_p E_p^+ \\
x_p^- &= \cos\phi_p E_p^- \\
u_k &= n_k \sec\phi_k \ldots (k = 0...p)
\end{aligned} \tag{2.3}
$$

leads to a set of $2p$ linear, homogenous equations for the $2p+1$ unknowns $x_0^+, x_k^\pm, (k = 1...p)$:

$$
\begin{aligned}
x_1^+ + x_1^- - x_0^+ &= 0 \\
u_1 x_1^+ - u_1 x_1^- - u_0 x_0^+ &= 0 \\
x_{k+1}^+ + x_{k+1}^- - e^{-iv_k d_k} x_k^+ - e^{iv_k d_k} x_k^- &= 0 \qquad (k = 1...p-1) \\
u_{k+1} x_{k+1}^+ - u_{k+1} x_{k+1}^- - u_k e^{-iv_k d_k} x_k^+ + u_k e^{iv_k d_k} x_k^- &= 0 \qquad (k = 1...p-1)
\end{aligned} \tag{2.4}
$$

Once solved for x_p^+ and x_p^- in terms of x_0^+, the transmissivity T_p and reflectivity R_p are given by:

$$T_p = \frac{n_0 \cos^2\phi_p}{n_p \cos^2\phi_0} \left| \frac{x_0^+}{x_p^+} \right|^2 \tag{2.5}$$

and

$$R_p = \left| \frac{x_p^-}{x_p^+} \right|^2 . \tag{2.6}$$

While the amplitude reflection coefficient will gain significance in a later section on ellipsometry

$$r_p = \frac{x_p^-}{x_p^+}. \tag{2.7}$$

2.2.2 Case II: Light Polarized Perpendicular to the Plane of Incidence

The procedure is similar to Case I, let $Z_k^\pm$ be the z-components of the electric field vectors traveling in positive and negative directions in medium k and $\beta_k^\pm$ the y-components of the corresponding magnetic field vectors:

$$Z_k^+ = E_k^+ \exp\frac{2\pi i}{\lambda}(n_k(x\cos\phi_k + y\sin\phi_k) - ct),$$

$$Z_k^- = E_k^- \exp\frac{2\pi i}{\lambda}(n_k(-x\cos\phi_k + y\sin\phi_k) - ct)$$

and

$$\beta_k^+ = -n_k \cos\phi_k E_k^+ \exp\frac{2\pi i}{\lambda}(n_k(x\cos\phi_k + y\sin\phi_k) - ct),$$

$$\beta_k^- = n_k \cos\phi_k E_k^- \exp\frac{2\pi i}{\lambda}(n_k(-x\cos\phi_k + y\sin\phi_k) - ct).$$

The boundary conditions are the same as before, so with the substitutions

$$\begin{aligned} y_k^+ &= E_k^+ \exp\{iv_k(d_{p-1} + d_{p-2} + ... + d_k)\} \ldots (k = 1...p-1) \\ y_k^- &= E_k^- \exp\{-iv_k(d_{p-1} + d_{p-2} + ... + d_k)\} \ldots (k = 1...p-1) \\ y_0^+ &= E_0^+ \exp\{iv_0(d_{p-1} + d_{p-2} + ... + d_1)\} \\ \omega_k &= n_k \cos\phi_k \ldots (k = 0...p) \end{aligned} \tag{2.8}$$

the set of equations for the y's is obtained:

$$\begin{aligned} y_1^+ + y_1^- - y_0^+ &= 0 \\ \omega_1 y_1^+ - \omega_1 y_1^- - \omega_0 y_0^+ &= 0 \\ y_{k+1}^+ + y_{k+1}^- - e^{-iv_k d_k} y_k^+ - e^{iv_k d_k} y_k^- &= 0 \qquad (k = 1...p-1) \\ \omega_{k+1} y_{k+1}^+ - \omega_{k+1} y_{k+1}^- - \omega_k e^{-iv_k d_k} y_k^+ + \omega_k e^{iv_k d_k} y_k^- &= 0 \qquad (k = 1...p-1) \end{aligned} \tag{2.9}$$

which has exactly the same form as Eqs. (2.4), if we exchange the x's to y's and the u's to ω's. Therefore the same solution will apply to both cases. The transmissivity T_p' and reflectivity R_p' are given by:

$$T_p' = \frac{n_0}{n_p}\left|\frac{y_0^+}{y_p^+}\right|^2 \tag{2.10}$$

and

$$R_p' = \left|\frac{y_p^-}{y_p^+}\right|^2. \tag{2.11}$$

2.2.3 Solution in Matrix Form

We will concentrate on solving Eqs. (2.4), since through renaming the variables, Eqs. (2.9) are solved as well. With a brief look at the last two lines of Eqs. (2.4), one finds that the solution is already in the form of an iteration algorithm, which becomes clearer if written in matrix form:

$$\begin{bmatrix} x_{k+1}^+ \\ x_{k+1}^- \end{bmatrix} = \frac{1}{2u_{k+1}} \begin{bmatrix} (u_{k+1} + u_k)e^{-iv_k d_k} & (u_{k+1} - u_k)e^{iv_k d_k} \\ (u_{k+1} - u_k)e^{-iv_k d_k} & (u_{k+1} + u_k)e^{iv_k d_k} \end{bmatrix} \cdot \begin{bmatrix} x_k^+ \\ x_k^- \end{bmatrix} \tag{2.12}$$

Where we can call the matrix of the right hand side M'_k, thus writing the full solution as

$$\begin{bmatrix} x_p^+ \\ x_p^- \end{bmatrix} = \frac{1}{2^p u_1 u_2 \cdots u_p} M'_{p-1} M'_{p-2} \cdots M'_1 \begin{bmatrix} (u_1 + u_0)x_0^+ \\ (u_1 - u_0)x_0^- \end{bmatrix} \tag{2.13}$$

Now let t_k be the Fresnel amplitude transmission coefficient for light passing from medium k to medium $k-1$ (i.e. in positive direction) and r_k the amplitude reflection coefficient for light coming from medium k that is reflected off the boundary with medium $k-1$ (i.e. $+ \to -$). These are for light polarized parallel to the plane of incidence:

$$t_k^p = \frac{2n_k \cos\phi_k}{n_k \cos\phi_{k-1} + n_{k-1} \cos\phi_k} \tag{2.14}$$

and

$$r_k^p = \frac{n_k \cos\phi_{k-1} - n_{k-1} \cos\phi_k}{n_k \cos\phi_{k-1} + n_{k-1} \cos\phi_k} \tag{2.15}$$

Now, with a look at Eq. (2.13) we can identify terms of the form $\frac{(u_k \pm u_{k-1})}{2u_k}$ as

$$\frac{(u_k + u_{k-1})}{2u_k} = \frac{\cos\phi_k}{\cos\phi_{k-1}} \frac{1}{t_k}$$

and

$$\frac{(u_k - u_{k-1})}{2u_k} = \frac{\cos\phi_k}{\cos\phi_{k-1}} \frac{r_k}{t_k}$$

Rewriting Eq. (2.13) in these terms yields

$$\begin{bmatrix} x_p^+ \\ x_p^- \end{bmatrix} = \frac{M_{p-1} \cdots M_2 \cdot M_1}{t_p \cdots t_2 t_1} \frac{\cos\phi_p}{\cos\phi_0} \cdot x_0^+ \cdot \begin{bmatrix} 1 \\ r_1 \end{bmatrix} \tag{2.16}$$

with the substitutions

$$\delta_k = v_k d_k$$

and

$$M_k = \begin{bmatrix} e^{-i\delta_k} & r_{k+1} e^{i\delta_k} \\ r_{k+1} e^{-i\delta_k} & e^{i\delta_k} \end{bmatrix}$$

Note that this is a very similar notation to that of standard text books, but it is not the same. With the use of Eqs. (2.7), (2.5), and (2.6) one can now calculate interesting quantities like the amplitude reflection coefficient of the stack or the intensity transmissivity and reflectivity:

$$r_{stack} = \frac{E_p^-}{E_p^+} = \frac{x_p^-}{x_p^+} = \frac{\begin{bmatrix} 0 \\ 1 \end{bmatrix} \left(M_{p-1} \cdots M_1 \cdot \begin{bmatrix} 1 \\ r_1 \end{bmatrix} \right)}{\begin{bmatrix} 1 \\ 0 \end{bmatrix} \left(M_{p-1} \cdots M_1 \cdot \begin{bmatrix} 1 \\ r_1 \end{bmatrix} \right)} \tag{2.17}$$

and

$$R = \left|\frac{x_p^-}{x_p^+}\right|^2 = \left|\frac{\begin{bmatrix} 0 \\ 1 \end{bmatrix}\left(M_{p-1}\cdots M_1 \cdot \begin{bmatrix} 1 \\ r_1 \end{bmatrix}\right)}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\left(M_{p-1}\cdots M_1 \cdot \begin{bmatrix} 1 \\ r_1 \end{bmatrix}\right)}\right|^2 \tag{2.18}$$

and

$$T = \frac{n_0\cos^2\phi_p}{n_p\cos^2\phi_0}\left|\frac{x_0^+}{x_p^+}\right|^2 = \frac{n_0}{n_p}\left|\frac{t_p\cdots t_1}{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\left(M_{p-1}\cdots M_1 \cdot \begin{bmatrix} 1 \\ r_1 \end{bmatrix}\right)}\right|^2 . \tag{2.19}$$

It is reassuring to see that r_{stack} reduces to r_1 if no layer is present. The above is for the case of light polarized parallel to the plane of incidence and the Fresnel coefficients r and t should be read accordingly. It can be shown that the solutions for the case of light polarized perpendicularly to the plane of incidence formally look exactly like Eqs. (2.17) (2.18) and (2.19) with their corresponding Fresnel coefficients, which are

$$t_k^s = \frac{2n_k\cos\phi_k}{n_k\cos\phi_k + n_{k-1}\cos\phi_{k-1}} \tag{2.20}$$

and

$$r_k^s = \frac{n_k\cos\phi_k - n_{k-1}\cos\phi_{k-1}}{n_k\cos\phi_k + n_{k-1}\cos\phi_{k-1}} \tag{2.21}$$

If a layer is absorbing, all the above formulae still apply if one substitutes the real index of refraction n by the complex index of refraction $n + ik$.

2.3 One and Two-Layer Antireflection Coatings

Applying Eq. (2.18) to the case of a single film on a substrate yields

$$R = \frac{r_1^2 + r_2^2 + 2r_1 r_2 \cos 2\delta_1}{1 + r_1^2 r_2^2 + 2r_1 r_2 \cos 2\delta_1}. \tag{2.22}$$

This result can also be obtained by summing up all the amplitude reflections that occur at the interfaces and considering the phase change while traversing the film (see e.g. [131]). The case of a single film is the only case that can be treated this way in a simple manner, as the procedure is needlessly complicated for $n \geq 2$. Nevertheless, it has been done for $n = 2$ and $n = 3$ and an iteration procedure has been found to deduce the case for k layers from the expression for $k - 1$ (see [132]). In terms of the refractive indices, at normal incidence Eq. (2.22) reads:

$$R = \frac{n^2(n_a - n_s)^2 - (n_a^2 - n^2)(n^2 - n_s^2)\sin^2(2\pi nd/\lambda)}{n^2(n_a + n_s)^2 - (n_a^2 - n^2)(n^2 - n_s^2)\sin^2(2\pi nd/\lambda)}. \tag{2.23}$$

We have renamed the variables, n, n_a, and n_s as the refractive indices of the layer, ambient and substrate, respectively. Here it is worth noting that if the layer has an optical thickness of half a wavelength $nd = \lambda/2$ the refractive index of the layer cancels out of the expression and the resulting reflectivity R has the form of a substrate without the layer. Thus the term 'absentee layer' for a half-wave film.

For the design of an antireflection coating, equating Eq. (2.23) to zero at some wavelength λ yields

$$n = \sqrt{n_s n_a} \qquad \text{and} \qquad 2\pi nd/\lambda = \pi/2, \quad 3\pi/2 \ldots \tag{2.24}$$

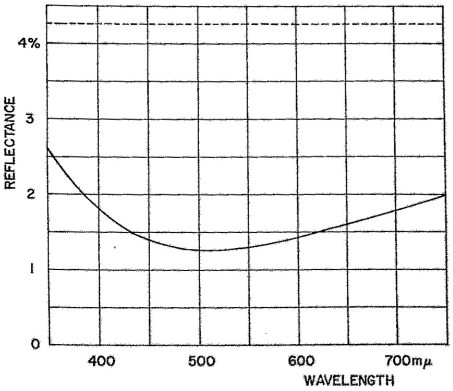
This constraint on the refractive index is a great drawback of single-layer antireflection coatings, since while the film thickness can easily be controlled, the refractive index is an inherent property of the material, and not tunable to match the requirement of a given substrate. Also, even if a material with the right index of refraction was available, the reflectivity increases differently towards the longer and shorter wavelengths, therefore introducing a 'color imbalance' in the transmitted light. See Fig. 2.2 for an example.

One way to overcome these problems is by introducing further layers. The equivalent of Eq. (2.23) for the case of two layers on a substrate looks considerably more tedious:

$$R = \frac{X}{1 + X} \tag{2.25}$$

with

$$\begin{aligned} X \;=\; & \frac{n_s}{4n_a}\Bigg\{\left[\left(\frac{n_a}{n_s} - 1\right)\cos\hat{\phi}_1 \cos\hat{\phi}_2 + \left(\frac{n_1}{n_2} - \frac{n_a n_2}{n_s n_1}\right)\sin\hat{\phi}_1 \sin\hat{\phi}_2\right]^2 \\ & + \left[\left(\frac{n_a}{n_1} - \frac{n_1}{n_s}\right)\sin\hat{\phi}_1 \cos\hat{\phi}_2 + \left(\frac{n_a}{n_2} - \frac{n_2}{n_s}\right)\cos\hat{\phi}_1 \sin\hat{\phi}_2\right]^2\Bigg\} \end{aligned} \tag{2.26}$$



Comparison of the reflectance of a single layer antireflection coating (solid curve) with the reflectance of the uncoated surface (dotted curve), $n_S = 1.52$, $n_M = 1.0$, $n_1 = 1.38$, $4n_1d_1 = 510$ nm.

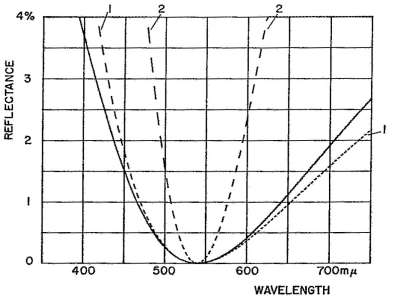
203

Figure 2.2 *Quarter-wave antireflection layer on glass. The index of refraction misses the square root condition by some small amount. Image taken from [132].*

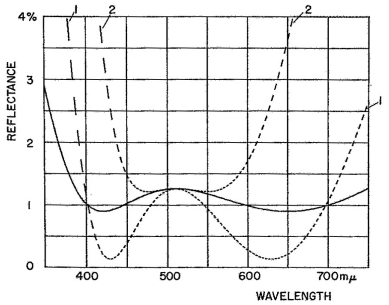
and

$$\hat{\phi}_i = 2\pi n_i d_i/\lambda \tag{2.27}$$

where this time, $i = 1$ is the top-most layer. It makes sense to simplify Eq. (2.25) by approximating $R \approx X$ for small R. Solving the problem of minimizing the reflection with two layers instead of one has the effect of introducing free parameters to expand the set of possible solutions. This means either to be able to reach $R = 0$ even though no material with a certain refractive index is available, or to expand the region of very low residual reflectivity to a greater wavelength region and making the spectral response more symmetric. These two approaches are called *V-type* and *W-type* coatings in literature [132], by the shape of their spectral responses, as can be seen in Fig. 2.3.



Reflectance of various two-layer V-type coatings on a substrate with index $n_S = 1.60$, $n_M = 1.0$;
solid curve: $n_1 = 1.38$, $4n_1d_1 = 540$ nm; $n_2 = 1.746$, $4n_2d_2 = 540$ nm;
dotted curve 1: $n_1 = 1.38$, $4n_1d_1 = 671.5$ nm; $n_2 = 2.30$, $4n_2d_2 = 113.2$ nm;
dotted curve 2: $n_1 = 1.818$, $4n_1d_1 = 540$ nm; $n_2 = 2.30$, $4n_2d_2 = 540$ nm.



Reflectance of various two-layer W-type coatings on a substrate with index $n_S = 1.52$, $n_M = 1.0$;
solid curve: $n_1 = 1.38$, $4n_1d_1 = 510$ nm; $n_2 = 1.60$, $4n_2d_2 = 1020$ nm;
dotted curve 1: $n_1 = 1.38$, $4n_1d_1 = 510$ nm; $n_2 = 2.00$, $4n_2d_2 = 1020$ nm;
dotted curve 2: $n_1 = 1.38$, $4n_1d_1 = 510$ nm; $n_2 = 2.50$, $4n_2d_2 = 1020$ nm.

Figure 2.3 *V- and W-type antireflection coatings on glass using different materials. The quarter-wave layer is the same as in Fig. 2.2. Images taken from [132].*

To obtain a V-type coating one sets $R = X = 0$, this yields

$$\tan^2 \hat{\phi}_1 = n_1^2 \frac{(n_a - n_s)(n_s n_a - n_2^2)}{(n_s n_1^2 - n_a n_s^2)(n_s n_a - n_1^2)} \tag{2.28}$$

$$\tan^2 \hat{\phi}_2 = n_2^2 \frac{(n_a - n_s)(n_s n_a - n_1^2)}{(n_s n_1^2 - n_a n_s^2)(n_s n_a - n_2^2)} \tag{2.29}$$

These relations allow for a wide range of n_1 and n_2 and yields the thicknesses.

To obtain a W-type coating, we insert a half-wave layer between the original (sub-optimal) quarter-wave layer and the substrate. This leaves the reflectivity at the design-wavelength unchanged ('absentee layer'), but, as long as the half-wave layer's refractive index fulfills certain loose conditions [132], it has the effect of turning the original minimum in reflectance into a maximum with two minima around it, thereby making the spectral response flatter. Setting $\phi_2 = 2\phi_1 = \pi/2$ in Eq. (2.26) yields

$$X = \frac{n_s}{4n_a} \left(\frac{n_a}{n_1} - \frac{n_1}{n_2} \right)^2 \tag{2.30}$$

and

$$R = \left(\frac{n_a n_s - n_1^2}{n_a n_s + n_1^2} \right)^2 \tag{2.31}$$

which is the result for one quarter-wave layer with n_1 as can be seen from Eq (2.23). Adding further layers provides even more degrees of freedom in design for the properties of antireflection coatings, or any kind of optical filters for that matter. Treatment of these is better done numerically with the aid of computers in minimizing or maximizing reflectivity or transmissivity at different wavelengths or ranges simultaneously, with or without absorption.

2.4 Production of Antireflection Coatings for QCL Facets

Production of AR coatings was done in our labs using reactive sputter deposition for which we developed a system to measure the laser emission quasi in-situ during the coating process.

The sputter deposition system consisted of a water-cooled target and sample table in a vacuum chamber. Pumping of the system was done in two stages using a rotary vane and a turbo pump. Incorporated into the chamber were gas inlets with mass flow controllers and an RF frequency generator for the plasma. The vacuum, plasma and gas flow were controlled with various electronic meters and valves. The target and sputter table served as electrodes for the RF electric field. The thickness of the deposited film was coarsely monitored in situ with a quartz oscillator that was placed next to the sample and coated along with it. Refer to Fig. 2.4 for a schematic setup and to Fig. 2.5 for a photograph of the inside of the chamber during deposition.

The residual pressure in the chamber prior to the inlet of the sputter gases was $2 \cdot 10^{-4}$ Pa. During sputtering the pumps were running, the 36 sccm/min of argon influx resulted in an equilibrium chamber pressure of 0.6 Pa. The resulting partial pressures for the gases at

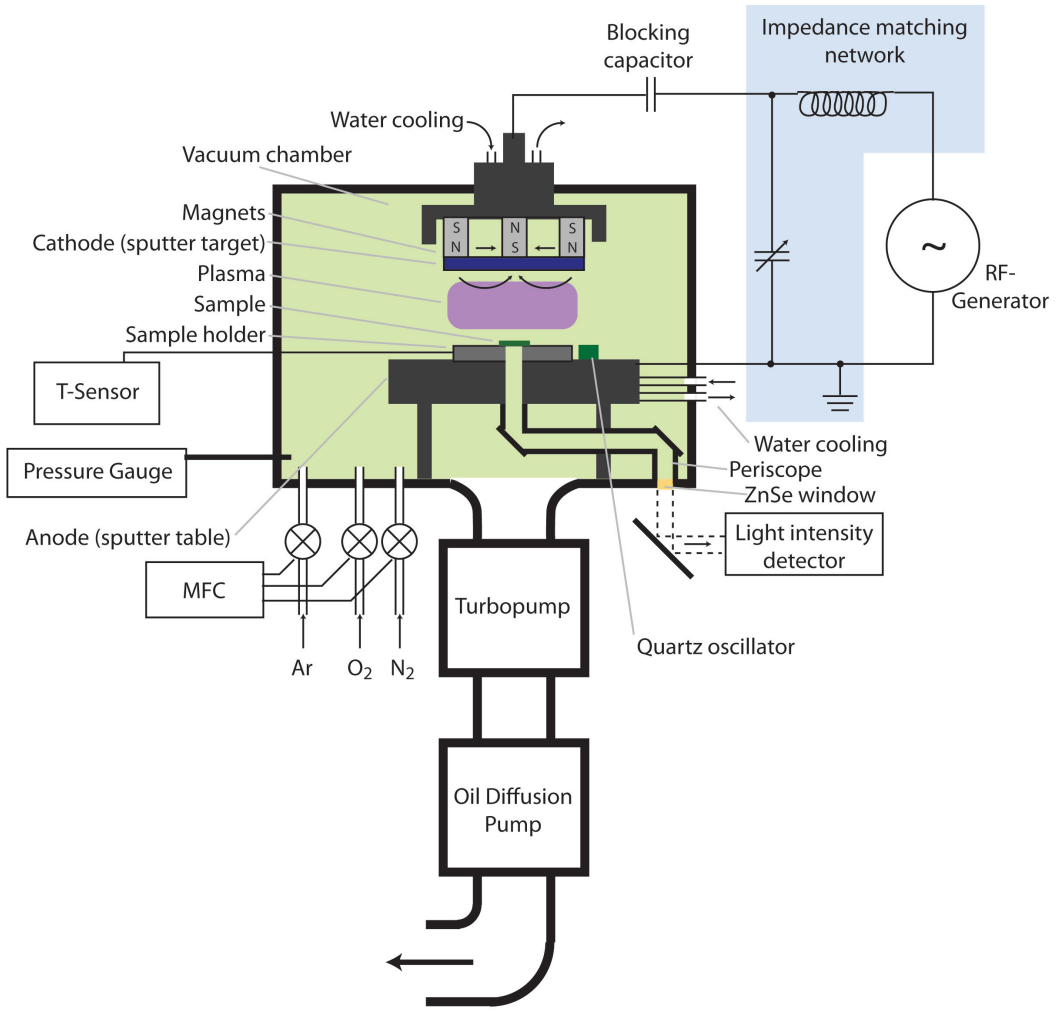


Figure 2.4 *Schematic setup of the sputter coater in use.*

different flow rates are given in Table 2.1. Due to the spatial variation of pressure in the chamber due to the highly non-equilibrium setup, there is a low degree of confidence in these values. The plasma had a power of 200-600 W depending on sample at an E-field frequency of 27.12 MHz and a DC acceleration voltage of 7-16 V depending on sample. The standing-wave ratio was 1.1 on all samples.

The available targets were 99.99 % pure Si, Al, Ti, TiO_2 and Al_2O_3 from KJ Lesker. To clean the targets, they were plasma oxidized for 60 s after insertion and the top layer was sputtered off during 20-min runs of non-reactive, pure-argon sputtering without a sample in the chamber. The target and substrate temperatures were not directly controlled.

The laser was mounted vertically inside a specially designed sample holder to be seen in Fig. 2.6. The cover plate of this sample holder contains a small slit, which allows only the laser facet underneath to be coated. The sample holder employs contacts for two coaxial cables, one to drive the laser inside the vacuum chamber without opening it and the other to perform in-situ temperature measurements during the coating process to avoid overheating of the sample. Figure 2.7 is a photograph of a QCL whose facets pointing towards the bottom

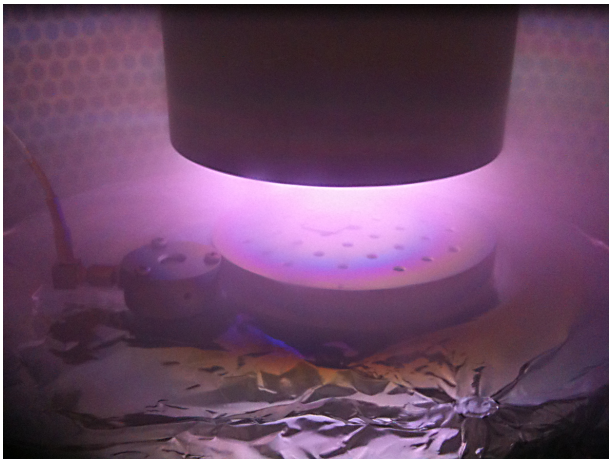


Figure 2.5 *Photograph of the deposition chamber during deposition.*

Gas flow in sccm/min	Partial pressure in Pa
4	$6 \cdot 10^{-2}$
2	$3 \cdot 10^{-2}$
1.8	$2.6 \cdot 10^{-2}$
1.6	$2.3 \cdot 10^{-2}$
1.4	$2.0 \cdot 10^{-2}$

Table 2.1 *Resultant partial pressures of O_2 and N_2 in sputter chamber for given influx while vacuum pumps are running.*

of the picture is coated with a thin layer of Y_2O_3.

The other, uncoated facet faces down onto the water-cooled sputter table. This table has a drilled hole of 10 mm diameter in its center. Inserted into this hole is a copper pipe, gold-coated on its inside, that collects and guides the light from the bottom facet of the QCL to a ZnSe window in the vacuum chamber. This very useful setup allows light from the QCL to be detected outside the sputter coater without the need to vent it. Before the coating process, a P-I curve is measured. The coating is then started and interrupted at given times to remeasure the P-I characteristic. This allows monitoring of when the layer has its optimal thickness and the sample can be removed. This is a quasi-in-situ measurement of the laser characteristics, since actual in-situ measurements are not possible, since the sputter plasma interferes with the radiation detector. The remainder of this chapter discusses the QCL's performance with a (partial) AR coating, as well as the advantages of performing in-situ monitoring along with the experimental procedure.

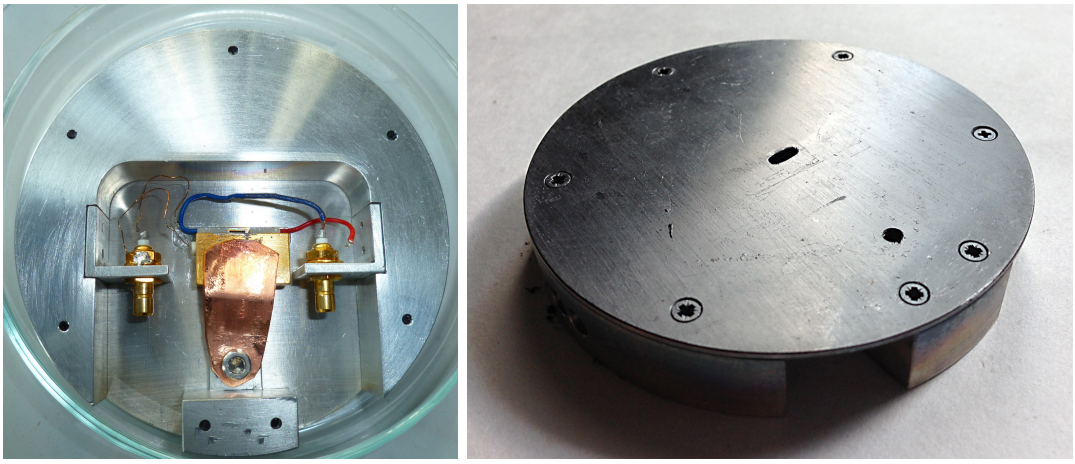


Figure 2.6 *Left: Bottom view of sample holder for coating laser facets. Clearly visible at the center is the brass C-mount that holds the vertically mounted QCL chip (seen facet-on as a short dark horizontal line) which is contacted by the blue and red wires leading to the right-hand SMB jack. The left-hand SMB jack connects to a small PT100 temperature sensor (white) inserted into the C-mount from the left. Right: Top view of sample holder for coating laser facets.*

2.5 Determining the reflectivity of an AR coating on a QCL facet

We have used Spectroscopic Mid-infrared Ellipsometry, described thoroughly in Appendix B, to find the dispersion relations of a number of materials and to determine their suitability for use in AR coatings. The results of our investigation of suitable materials has been published earlier [192]. We have also used Ellipsometry to calibrate the deposition system. The informations derived were then used to design the coatings with the use of a numerical program given in Appendix C.3 that optimizes the layer thicknesses according to the Matrix Method of the previous Section.

Since the performance of an AR-coated QCL is extremely sensitive to the residual reflectivity, it is not sufficient to rely on the measured optical properties and calibration, but instead we have also developed an in-situ monitoring system during deposition of the coatings. This measures the laser's emission as a function pump current inside the deposition chamber and during the coating process.

Although the ability to determine the laser output during deposition is the most direct way to measure the quality of the coating, especially in long QCLs, its interpretation is not entirely straight-forward.

When attempting to determine the residual reflectivity of a given laser facet antireflection (AR) coating (in-situ or ex-situ), or even when designing any coating to alter a laser's properties according to some requirement, it needs to be clear how given facet reflectivities alter the output of light for a given laser. The canonical formulas dealing with the subject, however, are only valid for facet reflectivities very close to 1. But a semiconductor laser, even without any coating, loses nearly three quarters of its photons every time they reach a facet. For a

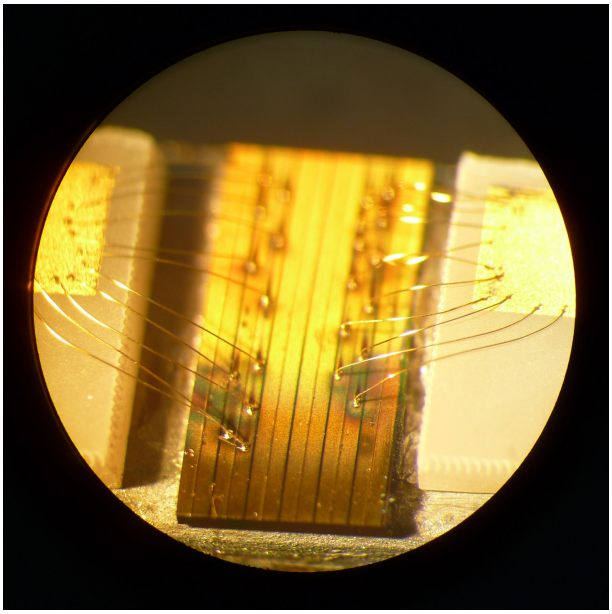


Figure 2.7 *View of the Y_2O_3-coated (lower edge) QCL chip mounted on its C-mount and contacted with bond wires.*

QCL with a good AR coating, the best approximation is that of an open resonator, and consequently it does not lase at all. The transition region between a laser and an open resonator that could – depending on how strongly it is pumped – act more like a light-emitting diode (LED) or a superluminescent diode (SLD) is not well-documented in the literature. Thus, we have developed and implemented a simple, numerically solvable model of a semiconductor laser resonator with arbitrary facet reflectivity, discussed in Section 2.5.4. This model, although crude, is versatile as it makes very few assumptions, and can also be used for different tasks such as optimizing facet reflectivities to achieve maximum output power or determining crucial laser parameters from fitting the model to measured Power-Current (P-I) curves.

The standard ways of determining the residual facet reflectivity after deposition of an AR coating, as found in the literature are described in the following along with the problems associated with them.

2.5.1 Reflectivity from threshold current

The threshold current density of a QCL according to [133] and [16], is

$$J_{\mathrm{th}} = \frac{\alpha_w + \alpha_{m1} + \alpha_{m2}}{\Gamma g}, \tag{2.32}$$

where $\alpha_{mi} = \frac{1}{L} \ln \frac{1}{|r_i|}$ are the mirror losses of facet i, Γ is the overlap of the laser mode with the active region, and g is the gain coefficient defined by $gJ = \gamma$. The ratio, $\bar{X}$, of the threshold currents, I_{th}, before and after coating results in

$$\bar{X} \equiv \frac{I_{\mathrm{th}}^{\mathrm{ac}}}{I_{\mathrm{th}}^{\mathrm{bc}}} = \frac{J_{\mathrm{th}}^{\mathrm{ac}}}{J_{\mathrm{th}}^{\mathrm{bc}}} = \frac{\alpha_w + \alpha_{m1} + \alpha_{m2}^{\mathrm{coated}}}{\alpha_w + 2\alpha_{m1}}, \tag{2.33}$$

where $I_{\text{th}}^{\text{bc}}$ and $I_{\text{th}}^{\text{ac}}$ are the threshold currents before and after coating, respectively. The mirror loss for the uncoated facets are $\alpha_{m1} = \alpha_{m2} = \frac{1}{L}\ln\frac{1}{|r_{\text{cleaved}}|}$ where $L = 0.6$ cm and $|r_{\text{cleaved}}| \approx 0.28$. The mirror loss of the coated facet is $\alpha_{m2}^{\text{coated}} = \frac{1}{L}\ln\frac{1}{|r_{\text{coated}}|}$. Equation (2.33) can be rearranged to give the reflectivity as a function of the ratio of threshold currents before and after coating to be

$$R_2^{\text{coated}} = \exp(-2L(\bar{X}(\alpha_w + 2\alpha_{m2}) - (\alpha_w + \alpha_{m2}))), \tag{2.34}$$

where $R_2^{\text{coated}} = |r_{\text{coated}}|^2$.

2.5.2 Reflectivity from the ratio of the slope efficiencies

A second method for determining the reflectivity of the coated facet is by comparing the slopes of curves measured at the two facets after coating. The power-current (P-I) characteristic of an ideal semiconductor laser above threshold is

$$P = \eta_d(I - I_{th})\frac{\hbar\omega}{e} = s(I - I_{th}), \tag{2.35}$$

where I is the current, I_{th} is the threshold current, e is the elementary charge, $s = \frac{\partial P}{\partial I}$ is the *slope efficiency*, and η_d is the *external differential quantum efficiency*

$$\eta_d^j = \eta_e^j \eta_i, \tag{2.36}$$

where η_i is the *internal quantum efficiency* and η_e^j is the *extraction efficiency* of facet j. But the extraction efficiency of one mirror is just the mirror loss of this mirror divided by the total loss

$$\eta_e^j = \frac{\alpha_{mj}}{\alpha_w + \alpha_{m1} + \alpha_{m2}}, \tag{2.37}$$

where j is 1 or 2. But since the internal quantum efficiency does not change when depositing a coating on a facet, the ratio of the slopes is

$$\bar{Y} = \frac{s^c}{s^{uc}} = \frac{\eta_e^c}{\eta_e^{uc}} = \frac{\alpha_m^c}{\alpha_m^{uc}}, \tag{2.38}$$

where s^c and s^{uc} are the slopes of the coated and uncoated facets respectively, η_e^c and η_e^{uc} are the extraction efficiencies of the respective facets and α_m^c and α_m^{uc} the mirror losses. But from this, the reflectivity of the coated facet can be determined to be

$$R_2^{\text{coated}} = \left(R_1^{\text{uncoated}}\right)^{s^c/s^{uc}}. \tag{2.39}$$

where $R_2^{\text{coated}} = |r_{\text{coated}}|^2$.

2.5.3 Problems with determining the reflectivity of a good AR coating

Figure 2.8 shows the P-I curves for an uncoated laser (HU2-0324, 6mm) and for both facets of the same laser after deposition of an AR coating onto one of them. The pulse duration was 100ns and the repetition rate was 87kHz.

The green dots represent the uncoated laser and look like a typical P-I curve for a QCL. It consists of a linear region with a clearly defined slope, as can be seen between approximately 4.6 and 7.2 A. On the left side it is bounded by the "knee", a curved piece of the graph where the slope increases with increasing current. On the right hand side, the slope starts to run into a plateau after which it would fall off and the output decreases with increasing current. This is the "roll-over" region. The threshold is defined as the intersection of the (extrapolated) linear part of the curve with the x-axis. The threshold of the green plot is clearly defined and found to be 4.51 A.

2.5.3.1 Reflectivity from threshold current

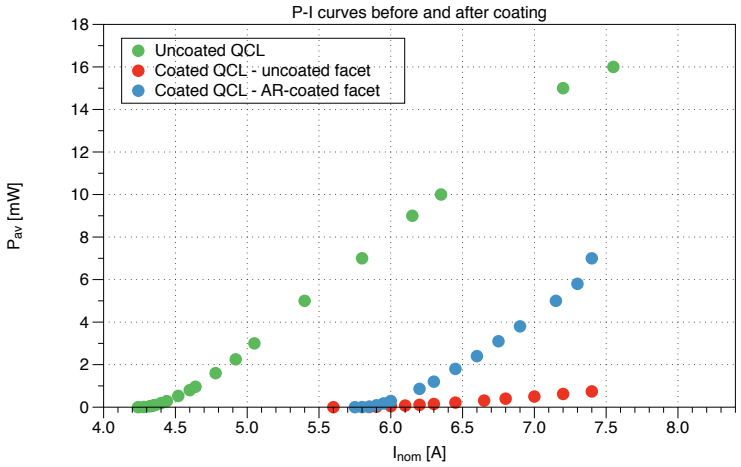


Figure 2.8 *P-I curves for the QCL before (green) and after (blue, red) deposition of an AR coating on one of the facets. Blue: P-I curve taken at the coated facet. Red: P-I curve taken at the uncoated facet.*

The problem with applying Eq. (2.34) is immediately apparent from Fig. 2.8. The blue plot represents the AR-coated side of the coated laser. But here, it is quite striking that the dots are curved *upwards* instead of downwards as would be expected from a rollover. Thus the rollover is not reached, and even a straight slope is not actually to be determined. Instead, the entire plot resembles the knee region of the green plot. Thus although it is very clear that the threshold current rises substantially, comparing the green dots with the blue or red dots, it is not clear which value to use specifically.

Figure 2.9 shows a few possible choices for the given situation. If we say the entire measured range is within the knee region of the laser's P-I curve and that we haven't in fact reached the straight slope yet, we could say the best possible choice for the slope (that is to extrapolate

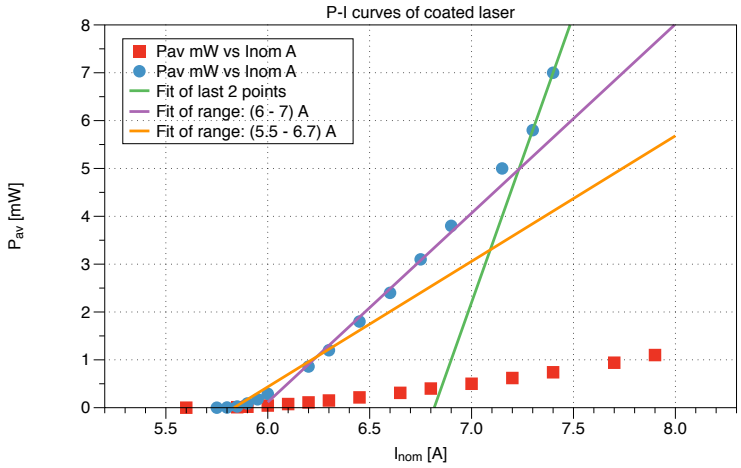


Figure 2.9 *Different choices for slopes of the curve of the coated facet in Fig. 2.8.*

to get the threshold) are the last two points. This is indicated by the green fit line and leads to a threshold current of 6.82 A. If we don't trust these last two values very much and decide to be more conservative in our estimate, we might choose the violet line as a fit of the most linear points. This would yield a threshold current of 5.98 A. If we want to be even more conservative, we might go as far as choosing the orange line that yields a threshold of 5.83 A. The three different choices of threshold currents would result in residual reflectivities for the AR coating of 0.7%, 2.5%, or 3.3%, respectively. The range between the largest and the smallest value is a factor of 5!

2.5.3.2 Reflectivity from slope efficiency

Choosing the purple and the gray lines in Fig. 2.10 as the slopes for the two (post-coating) plots in Fig. 2.8, 3.95 mW/A and 0.46 mW/A, according to Eq. (2.39), the residual reflectivity becomes $R_2^{\text{coated}} = 1.8 \times 10^{-5}$, which is 3 orders of magnitude smaller than the smallest number determined with the method of comparing thresholds!

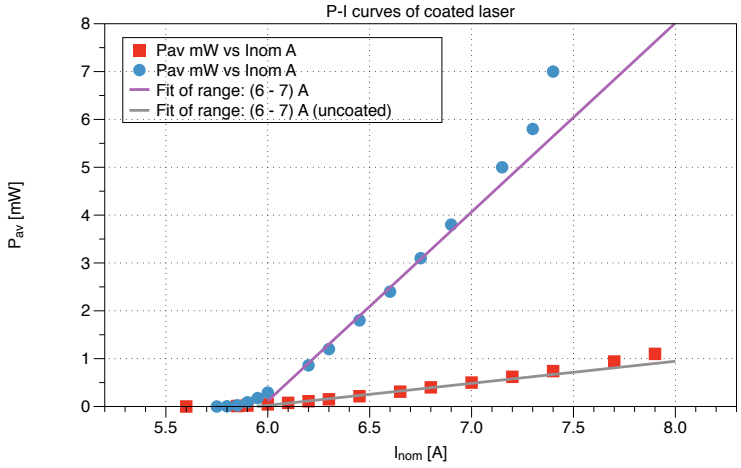


Figure 2.10 *Choices for the slopes of the coated and uncoated facets of the coated QCL from Fig. 2.8*

This discrepancy illustrates the problems associated with the assumptions that the AR-coated QCL still behaves well like a laser. So for $R_1 \approx 0$, the canonical approximations made for a laser are invalid. When feedback from one facet is nearly eliminated, the QCL behaves like a light emitting diode (LED), or for higher pump rates, as a super-luminescent diode (SLD). Thus, neither the threshold, nor the slope efficiency of the laser are clearly defined resulting in vastly different values for calculated residual reflectivities depending on how exactly one interprets the experimental values. But even for reflectivities of ≈ 0.28 the canonical form is not a good approximation and needs to be generalized as will become clear in the next sections.

2.5.4 Model for photon flux density distribution in saturated amplifier with $R \ll 1$

To generalize the usual form of laser behavior discussed in Section 1.4.1, we start at a more basic level than is given by Eq. (1.2) and develop a model without making some of the simplifications that lead to Eq. (1.2). This model uses elements of the theory of fiber lasers and amplifiers, the expressions in use can be found in many text books, for instance [133]. However, to keep to a level of complexity well-fit for our purposes, any phase-dependence of the electric field is omitted in this discussion, and photon flux densities that are proportional to E^2 are considered instead. However, the phase information can be added back in to arrive at expressions similar to Eq. (1.6), but the resulting expressions are far more tedious to write down or to compute. A more detailed discussion of this will be given in Sections 2.5.7 and 2.5.8.

Consider a general laser with an upper and a lower laser level. The probability of absorption of a photon by an electron in the lower level per second is

$$W = \phi\sigma(\nu), \tag{2.40}$$

where ϕ is the photon flux density (in photons per unit area per unit time) and σ is the transition cross section

$$\sigma(\nu) = \frac{\lambda^2}{8\pi t_{\mathrm{sp}} c} l(\nu), \tag{2.41}$$

where $\nu = 1/\lambda$ is the wavenumber of the light, λ is the wavelength, t_{sp} is the spontaneous emission lifetime of the transition and $l(\nu)$ is the normalized lineshape function of the transition (of width $\Delta\nu$ and peak value of $\frac{2}{\pi\Delta\nu}$). The probability for stimulated emission by an electron in the upper laser level is the same as Eq. (2.40).

If N_1 is the density of electrons in the lower state (in electrons per unit volume) and N_2 is the density of electrons in the upper state, then the rate of photons generated through stimulated emission (per unit volume) is $N_2 W$ and the rate of photons absorbed is $N_1 W$. We call $N = N_2 - N_1$ the population difference and if $N > 0$, there is population inversion, i.e. more electrons in the upper state. Thus if $N > 0$, photons traveling in the z direction

will release more photons traveling in the z direction through stimulated emission than are absorbed and – if spontaneous emission is ignored for now – the net rate of gained photons in an infinitesimal volume is just the rate of photon emission minus the rate of photon absorption.

$$\frac{d\phi}{dz} = (N_2 - N_1)W = NW = N\phi(z)\sigma(\nu) = \phi(z)\gamma(\nu), \tag{2.42}$$

where we have introduced the *gain*

$$\gamma(\nu) = N\sigma(\nu), \tag{2.43}$$

where the unit of the gain is 1/cm. The light intensity is related to the photon flux density through

$$\hat{I}(\nu) = hc\nu\phi(\nu), \tag{2.44}$$

where h is Planck's constant and c is the speed of light in vacuum.

It is clear that every absorption event reduces N_1 and ϕ by one and increases N_2 by one. On the other hand, every stimulated emission event reduces N_2 by one and increases N_1 and ϕ. So there will be a balance between N_1, N_2, ϕ, at every given pumping rate $\hat{R}$. Calculation of this balance is straight-forward with the use of the rate equations

$$\frac{dN_2}{dt} = \hat{R}_2 - \frac{N_2}{\tau_2} - N_2W + N_1W \tag{2.45}$$

and

$$\frac{dN_1}{dt} = -\hat{R}_1 - \frac{N_1}{\tau_1} + \frac{N_2}{\tau_{21}} + N_2W - N_1W, \tag{2.46}$$

where $\hat{R}_1$ and $\hat{R}_2$ are the pumping rates into levels 1 and 2, τ_1 and τ_2 are the overall lifetimes of carriers in levels 1 and 2 due to both radiative and non-radiative decay and $1/\tau_{21}$ is the rate of transitions from level 2 to 1. Figure 2.11 shows a sketch of the situation.

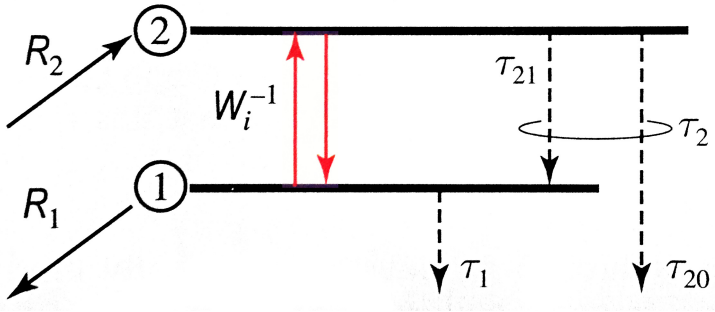


Figure 2.11 *The population densities N_1 and N_2 ($cm^{-3}s^{-1}$) of carriers in energy levels 1 and 2 are determined by three processes: decay (at rates $1/\tau_1$ and $1/\tau_2$, respectively, which includes the effects of spontaneous emission), depumping and pumping (at rates $\hat{R}_1$ and $\hat{R}_2$, respectively), and absorption and stimulated emission (at rate W_i with corresponding time constant $1/W_i$). Figure taken from [133]*

These equations can be solved for the steady-state ($\frac{dN_2}{dt} = 0$ and $\frac{dN_1}{dt} = 0$), using $N = N_2 - N_1$, to yield

$$N = \frac{N_0}{1 + \tau_s W} \tag{2.47}$$

where

$$\tau_s = \tau_2 + \tau_1(1 - \frac{\tau_2}{\tau_{21}}) \tag{2.48}$$

and

$$N_0 = \hat{R}_2\tau_2(1 - \frac{\tau_1}{\tau_{21}}) + \hat{R}_1\tau_1. \tag{2.49}$$

N_0 is the steady-state solution to N in the absence of radiation ($W = 0$).

Substituting Eq. (2.47) and Eq. (2.40) into Eq. (2.43) gives

$$\gamma(\nu) = \frac{\gamma_0(\nu)}{1 + \phi/\phi_s(\nu)} \tag{2.50}$$

where

$$\gamma_0(\nu) = N_0\sigma(\nu) \tag{2.51}$$

is the gain in the absence of amplifier radiation and $\phi_s = \frac{1}{\tau_s\sigma(\nu)}$ is called the *saturation photon flux density*. $\gamma_0(\nu)$ is the gain at the very beginning of the pumping process, where population inversion has been established through rapid pumping but not enough light has been produced by spontaneous emission yet to seed stimulated emission. $\gamma_0(\nu)$ is therefore called the *small-signal gain*. As is clear from Eq. (2.50), as the photon flux ϕ inside the resonator grows, the overall gain γ decreases, thus γ is called the *saturated* gain. When $\phi = \phi_s$, the overall gain has decreased to half its value for the empty resonator, γ_0.

From Eq. (2.50) it is clear that the gain is a function of photon flux, thus Eq. (2.42) becomes

$$\frac{d\phi}{dz} = \phi(z)\frac{\gamma_0(\nu)}{1 + \phi(z)/\phi_s(\nu)} \tag{2.52}$$

and thus the gain is also a function of z, thus, the solution to Eq. (2.50) is not a simple exponential function as it may seem from Eq. (2.43) due to gain saturation. Even in the regime of very little light it is not, since Eq. (2.50) still ignores the presence of spontaneous emission, which in the limiting case of very little stimulated light is not a good approximation. But this will be resolved soon.

Absorption is not the only way to lose photons, but there are other loss mechanisms. The first one to be treated is the scattering of photons out of the system. Introducing the phenomenological distributed waveguide loss α_w (in units 1/cm), the propagation equation of the photon flux density becomes

$$\frac{d\phi}{dz} = \phi(z)(\gamma(\nu, z) - \alpha_w). \tag{2.53}$$

Equation (2.53) only considers the photons traveling in the positive z-direction, but naturally there are also photons traveling in the negative z-direction. Making the obvious generalization, Eq. (2.53) splits up into two equations

$$\frac{d\phi_1}{dz} = \phi_1(z)(\gamma(\nu, z) - \alpha_w) \tag{2.54}$$

and

$$\frac{d\phi_2}{dz} = -\phi_2(z)(\gamma(\nu, z) - \alpha_w), \tag{2.55}$$

where $\phi_1(z)$ now denotes the photons traveling in the positive z direction and $\phi_2(z)$ denotes the photons traveling in the reverse direction. These equations are coupled since $\gamma(\nu, z)$ is saturated by photons traveling in both directions, thus Eq. (2.50) becomes

$$\gamma(\nu, z) = \frac{\gamma_0(\nu)}{1 + \frac{\phi_1(z)+\phi_2(z)}{\phi_s(\nu)}} \tag{2.56}$$

and with this, Eq. (2.54) and (2.55) become

$$\frac{d\phi_1}{dz} = \phi_1(z) \left(\frac{\gamma_0(\nu)}{1 + \frac{\phi_1(z)+\phi_2(z)}{\phi_s(\nu)}} - \alpha_w \right) \tag{2.57}$$

and

$$\frac{d\phi_2}{dz} = -\phi_2(z) \left(\frac{\gamma_0(\nu)}{1 + \frac{(\phi_1(z)+\phi_2(z)}{\phi_s(\nu)}} - \alpha_w \right). \tag{2.58}$$

The next step is to insert a phenomenological term for the spontaneous emission, ξ_{sp}, whose role will be more closely examined later. With this, Eq. (2.57) and (2.58) become

$$\frac{d\phi_1}{dz} = \phi_1(z) \left(\frac{\gamma_0(\nu)}{1 + \frac{\phi_1(z)+\phi_2(z)}{\phi_s(\nu)}} - \alpha_w \right) + \xi_{sp} \tag{2.59}$$

and

$$\frac{d\phi_2}{dz} = -\phi_2(z) \left(\frac{\gamma_0(\nu)}{1 + \frac{(\phi_1(z)+\phi_2(z)}{\phi_s(\nu)}} - \alpha_w \right) - \xi_{sp}. \tag{2.60}$$

To solve these equations, a set of boundary conditions is required. These are connected to the last loss mechanism for photons. At the (partially reflective) facets at each end of the waveguide, the light impinging on them generally splits up into two parts, the light that is transmitted and lost from the resonator, this is the useful light, and a part that is reflected back into the resonator, seeding the amplifier in the reverse direction. Thus the boundary conditions are

$$\phi_2(l) = R_2\phi_1(l) \tag{2.61}$$

and

$$\phi_1(0) = R_1\phi_2(0), \tag{2.62}$$

where l is the length of the wave guide (resonator), R_1 is the reflectivity of the facet at $z = 0$, and R_2 is the reflectivity of the facet at $z = l$.

For a given set of parameters α_w, $\phi_s(\nu)$, $\gamma_0(\nu)$, and ξ_{sp}, the boundary value problem consisting of Eq. (2.59)-(2.62) can easily be solved numerically, for instance with use of the Shooting Method, yielding self-consistent solutions for the photon density distributions $\phi_1(z)$ and $\phi_1(z)$ for $z \in (0, l)$. These solutions can now be related to observable output powers P_1 and P_2 from facets 1 and 2, respectively, through

$$P_1 = \phi_2(0)(1 - R_1)A\hbar\omega \tag{2.63}$$

and

$$P_2 = \phi_1(l)(1 - R_2)A\hbar\omega, \tag{2.64}$$

where A is the facet area and $\hbar\omega$ is the energy of a photon. At this point, it is not yet clear, how the laser parameters α_w, $\phi_s(\nu)$, $\gamma_0(\nu)$, and ξ_{sp} are to be determined for a realistic laser. This will be discussed in the following.

A laser can be viewed simply as a device that produces light when supplied with energy through pumping. In most semiconductor lasers, pumping is done by applying a voltage and running an electric current through the structure. The laser parameters now determine the response of the device to this pump current I. Thus it is a viable assumption that (most of) the laser parameters can be derived from the P-I curve, i.e. the response of output power P to pump current I.

To introduce the pump current into the problem, the laser parameters have to be brought in connection with it. Now while α_w, $\phi_s(\nu)$ are quite obviously, with sufficient accuracy, invariant to current changes, reasonable *ansatzes* have to be made for $\gamma_0(\nu)$ and ξ_{sp}. These are the following. In interband lasers, the small-signal gain coefficient $\gamma_0(\nu)$ is proportional to the pump rate $\hat{R}$. The pump rate, however, over a wide range is proportional to the pump current. Although this range is far smaller for QCLs than it is for interband lasers, it is reasonable to state that even in QCLs, over any range of interest (where light is actually emitted), this is a good approximation. Thus

$$\gamma_0(\nu) \approx \bar{g}I \tag{2.65}$$

where $\bar{g}$ is a new parameter, the gain coefficient that is constant with respect to current.

The spontaneous emission factor ξ_{sp} is proportional to the population density of electrons in the upper laser state, N_2. The first assumption to be made now is that the lower laser state mainly gets filled by electrons from the upper laser state, thus ignoring thermal backfilling and accidental non-radiative relaxation into the lower state from any state other than the

upper laser state (e.g. from the continuum). The second assumption is that the fraction of electrons from the upper laser state that reaches the lower laser state is constant with respect to current. If these assumptions are satisfactorily met, we can say the population density of the lower laser state N_1 is proportional to N_2. But then $N_2 \propto N_2 - N_1 = N$. But since $N \propto \gamma$, according to Eq. (2.43), $N \propto I$ if we disregard gain saturation, which is not a problematic assumption, given that spontaneous emission is most important in ranges where there is little light and thus the gain is hardly saturated. But this means

$$\xi_{sp} \approx \hat{\xi}_{sp} I, \tag{2.66}$$

where $\hat{\xi}_{sp}$ now is a parameter that is constant with respect to current.

These parameters can now be inserted into Eqs. (2.59) and (2.60) to give the full problem in terms of the pump current I as

$$\frac{d\phi_1}{dz} = \phi_1(z) \left(\frac{\bar{g} I}{1 + \frac{\phi_1(z) + \phi_2(z)}{\phi_s(\nu)}} - \alpha_w \right) + \hat{\xi}_{sp} I \tag{2.67}$$

and

$$\frac{d\phi_2}{dz} = -\phi_2(z) \left(\frac{\bar{g} I}{1 + \frac{(\phi_1(z) + \phi_2(z)}{\phi_s(\nu)}} - \alpha_w \right) - \hat{\xi}_{sp} I. \tag{2.68}$$

These equations in conjunction with an experimental P-I curve ($P_1(I)$ and $P_2(I)$) determine the boundary value problems of the photon flux densities for all currents I. Now for any *one* given current, the boundary value problem is highly under-determined in terms of the parameters $\bar{g}$, $\phi_s(\nu)$, α_w, $\hat{\xi}_{sp}$, R_1, and R_2, consequently there is a manyfold in the space spanned by these parameters that satisfies each boundary value problem. However, the manifold of parameter sets to solve the problems for *all* measured pairs of P-I values of the curve is much smaller and can, with some carefully chosen external information, be reduced to a single solution. This is discussed next.

As discussed earlier, a typical P-I curve for a laser is a linear function of the current I, intersecting the I-axis at the threshold current I_{th}. When measured carefully, near the threshold one finds a small bend, the "knee" of the curve. Two pieces of information can be determined from the linear function, the threshold and the slope, and a third from the curvature of the knee. This can fix three parameters. But since there are 6 parameters to be determined, the problem is obviously still under-determined. Thus the best way to determine all parameters is by first fitting the P-I curve of the uncoated laser, since then two parameters, the as-cleaved reflectivities of the laser facets, are known to be $R_1 = R_2 \approx 0.28$. If only one more parameter is determined through external measurement – e.g. α_w that can be calculated by measuring the threshold current as a function of laser length – the problem is fully determined and the remaining parameters, $\bar{g}$, $\phi_s(\nu)$, $\hat{\xi}_{sp}$ will be stably found through fitting.

With the exception of R_1, none of the parameters change through applying an AR coating

on facet 1. Thus, with the other parameters given, the P-I curves (both facets) of the coated laser can now be fitted with only R_1 open. The value R_1 fits to is the most accurate value we can establish for the quality of the coating. Especially when $R_1 \approx 0$, it is important to determine the residual reflectivity through this method due to the problems stated in Section 2.5.3.

2.5.5 Analysis of the model

This section briefly discusses the shape of the solutions to the model. Figure 2.12 shows the simulated internal photon flux density distribution modeled with the parameters $L = 0.6$ cm, $\hat{\xi}_{sp} = 0.0005 \times 10^{24}$ photons $(\text{cm}^3\ \text{s})^{-1}$, $\bar{g} = 1.384\ (\text{cmA})^{-1}$, $\alpha_w = 4\ \text{cm}^{-1}$, $\phi_s = 85.4 \times 10^{24}$ photons $(\text{cm}^2\ \text{s})^{-1}$. The blue curves represent the photons traveling to the right, $\phi_1(z)$, the green curves represent the photons traveling to the left, $\phi_2(z)$, while the red curve is the sum of both, $\phi_1(z) + \phi_2(z)$.

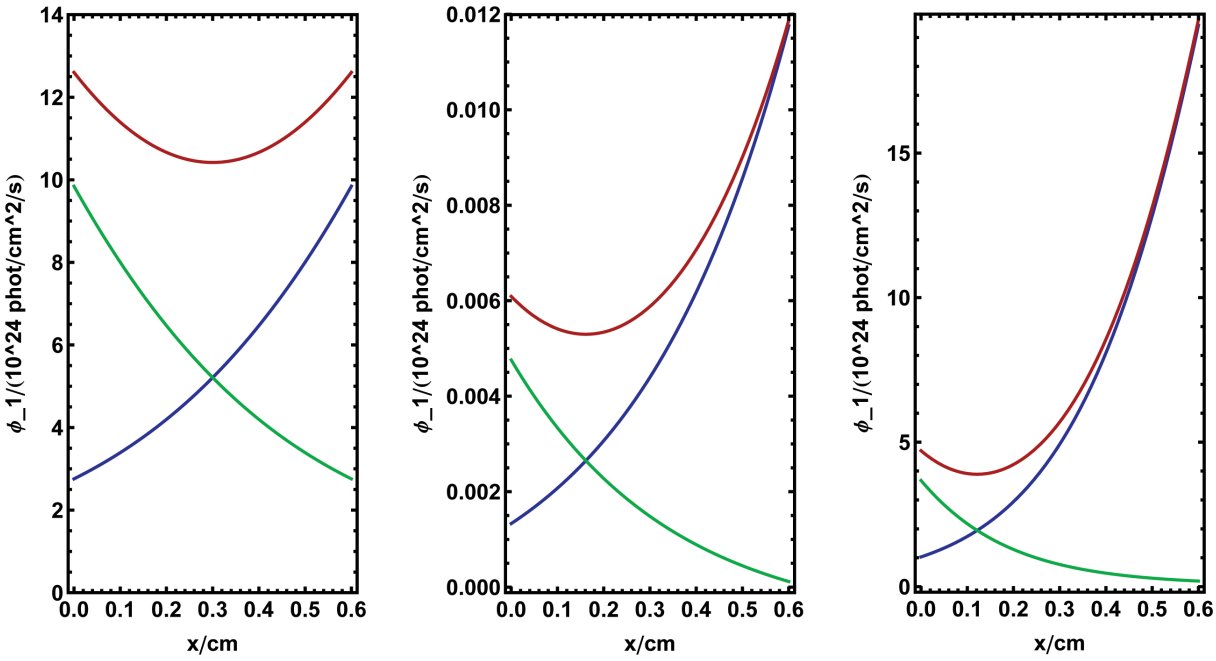


Figure 2.12 *Simulated internal photon flux density distribution modeled with the parameters* $L = 0.6$ *cm,* $\hat{\xi}_{sp} = 0.0005 \times 10^{24}$ *photons* $(cm^3\ s)^{-1}$, $\bar{g} = 1.384\ (cmA)^{-1}$, $\alpha_w = 4\ cm^{-1}$, $\phi_s = 85.4 \times 10^{24}$ *photons* $(cm^2\ s)^{-1}$. *Blue curve:* $\phi_1(z)$; *Green curve:* $\phi_2(z)$; *Red curve:* $\phi_1(z) + \phi_2(z)$. *Left plot: Uncoated, symmetric resonator with* $R_1 = R_2 = 28\%$ *at* $I = 5A$. *Center plot: Coated, asymmetric resonator with* $R_1 = 28\%$ *and* $R_2 = 1\%$ *at* $I = 5A$. *Right plot: Coated, asymmetric resonator with* $R_1 = 28\%$ *and* $R_2 = 1\%$ *at* $I = 7\ A$.

The left plot shows the uncoated, symmetric resonator with $R_1 = R_2 = 28\%$ at $I = 5$ A. The green and blue branches are clearly curved upward in their respective directions of propagation and have similarities to exponential functions. The red line has a minimum in the center of the resonator, and since it is this combined photon density that saturates the gain, the gain has a maximum at the center. Note how at the facets, at $z = 0$ and $z = 0.6$ cm, due to the reflectivity of 0.28, the reflected photon flux has a value of 0.28 times the outgoing, which are

the boundary conditions of the problem. In these areas, where the amplitude of the outgoing wave has a greater amplitude than the returning wave, the combined wave (red) nearly has the form of an outward traveling wave. Only towards the center, where $\phi_1 = \phi_2$, does the wave take the shape of a standing wave.

The center plot shows the case where the right facet is AR-coated and has a residual reflectivity of $R_2 = 1\%$ and is driven at the same current. Thus, at $z = 0.6$ cm, the green curve is a hundred times smaller than the blue curve and the resonator is not symmetric. It is interesting that although most light is *lost* to the right, the steady state demands that the internal photon density is highest at this end. Note also that the absolute value of the photon flux is three orders of magnitude smaller than for the left plot. This is because due to the increased losses from the right facet, the lasing threshold is not reached yet and what can be seen here is only amplified spontaneous emission. The right plot shows the same coated laser as the central plot, but driven above threshold at $I = 7$ A to reach approximately the same magnitude of photon flux as the case plotted on the left. Note how the blue and green curves in the right plot have a much greater curvature than the equivalent curves in the central plot. This is due to the fact that due to stronger pumping there is much greater gain (exponential character), while the curves in the central plot are more strongly dominated by spontaneous emission (linear character).

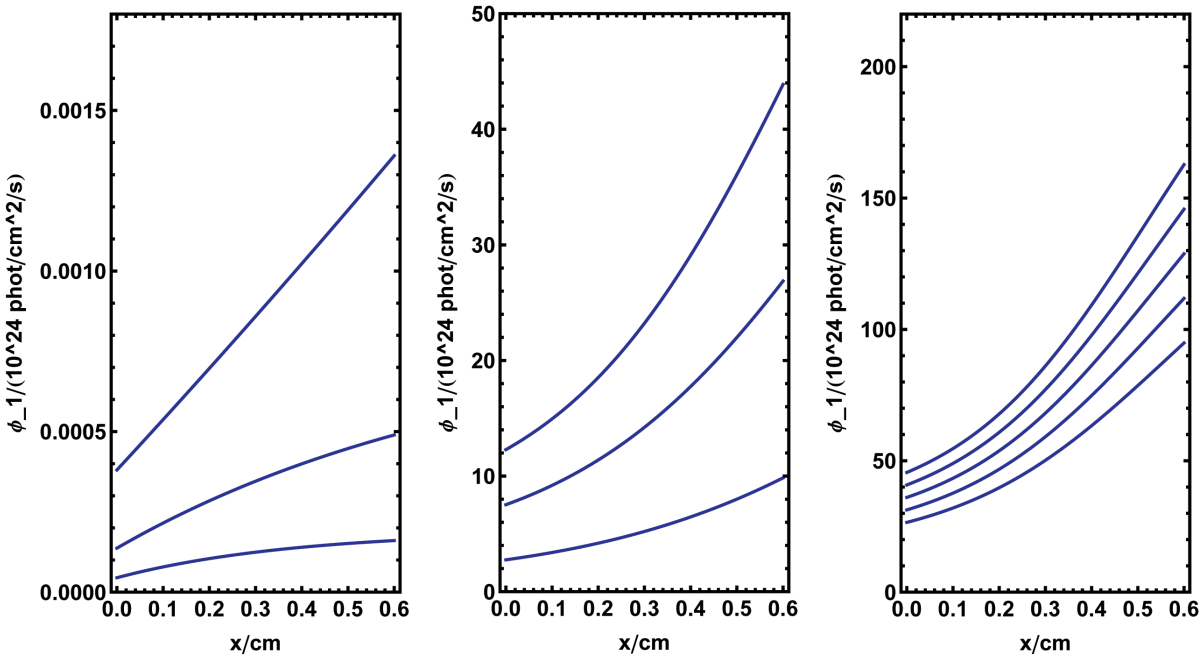


Figure 2.13 *Simulated internal photon flux density distribution, $\phi_2(z)$, for the uncoated, symmetric resonator with $R_1 = R_2 = 28\%$, modeled with the parameters $L = 0.6$ cm, $\hat{\xi}_{sp} = 0.0005 \times 10^{24}$ photons $(cm^3\ s)^{-1}$, $\bar{g} = 1.384\ (cmA)^{-1}$, $\alpha_w = 4\ cm^{-1}$, $\phi_s = 85.4{\times}10^{24}$ photons $(cm^2\ s)^{-1}$. Left plot (upwards): $I = (1, 2, 3)$ A. Center plot (upwards): $I = (5, 6, 7)A$. Right plot (upwards): $I = (10, 11, 12, 13, 14)A$. The left-traveling photons were not plotted, since in a symmetric resonator they are just the mirror images of the plotted curves.*

Figure 2.13 shows the internal photon flux density distribution of the same uncoated laser as in Fig. 2.12 a), modeled with the same parameters, but driven at various currents. The

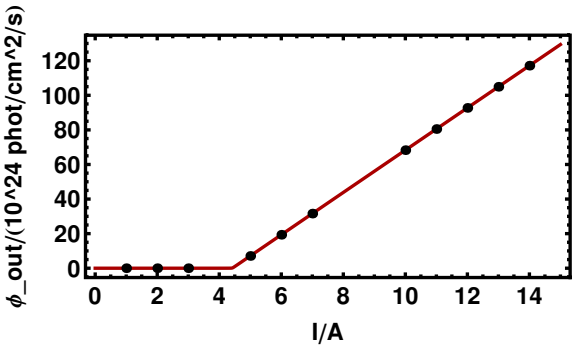


Figure 2.14 *P-I curve associated with the plotted photon densities of Fig. 2.13. Dots mark the currents at which the photon densities are plotted in Fig. 2.13. For better readability, the power is plotted in units of photon flux density, but this can be converted to actual power with the use of Eq. (2.63).*

photons traveling to the left and the sum $\phi_1(z) + \phi_2(z)$ are omitted, since for a symmetric resonator $\phi_2(z) = \phi_1(L - z)$, and the green and red curves are redundant. The left plot shows current values of 1, 2, and 3 A (lowest curve first). These are all below the lasing threshold as can be seen in Fig. 2.14, which shows the corresponding simulated P-I curve, where the current values of the various plots in Fig. 2.13 are marked (the first three dots that lie below the threshold kink are the values of the curves in the left plot and so on). For better comparability with Fig. 2.13, the power plotted Fig 2.14 is in units of photon flux density, but this can be converted to actual power with the use of Eq. (2.63). The curve at 3 A is very nearly linear. This is because at 3 A the gain γ almost exactly equals the distributed loss α_w. This is the so-called *transparent resonator* case, where there is neither gain nor loss and the only light propagating is the unobstructed (but accumulating) spontaneous emission. The two lower-lying curves are bent downwards in the direction of propagation, meaning loss per unit distance is greater than gain, thus the accumulating spontaneous emission is partially scattered. The curves in the central plot are taken at 5, 6, and 7 A, which all lie above threshold. It can be seen that the absolute photon densities are 4-5 orders of magnitude higher than in the left plot and that towards higher pump currents the curves get more curved, i.e. more exponential in character. The right plot shows pump currents of 10, 11, 12, 13, and 14 A, which is approximately three times the threshold current. Here gain saturation start to play a role, since the sum of the photon flux densities reach up to 200 scale units, while the saturation photon flux is 85. Thus the curves overall lie closer together since stronger pumping is partially compensated by stronger saturation.

2.5.6 Application of the model to determine the residual facet reflectivity

The model was employed to analyze, among others, an AR coating deposited on the QCL with the sample number HU2-0324. This coating was a 605-nm, quarter-wave, single-layer AR-coating of Y_2O_3. We have characterized the Y_2O_3 layers deposited with our sputter deposition system to be transparent at this wavelength and to have a refractive index of 1.843.

For the analysis, the P-I curves were measured before and after the coating process, see Fig. 2.8 as well as Fig. 2.15, and simulated using the program displayed in Appendix C.1. The procedure was as follows. As discussed earlier, one parameter needs to be fixed with an external measurement, because of the under-determination of the model information. We chose this parameter to be α_w, since it is easy to be determined from measuring the threshold current of an uncoated laser as a function of stripe length. For this we used a second stripe from the same processed wafer and found $\alpha_w = 4$ cm^{-1}. Further, we fixed the parameters $R_1 = R_2 = 0.28$ and $L = 0.6$ cm. With these parameters, we simulated the uncoated QCL's P-I curve first, which resulted in the fitted parameters $\phi_s = 108 \times 10^{24}$ photons (cm^2 s)$^{-1}$, $\hat{\xi}_{sp} = 0.000676 \times 10^{24}$ photons (cm^3 s)$^{-1}$, and $\bar{g} = 1.404$ (cmA)$^{-1}$. The result of this simulation is plotted in Fig. 2.15 (green line) along with the measured values (green dots) in a logarithmic plot for better readability.

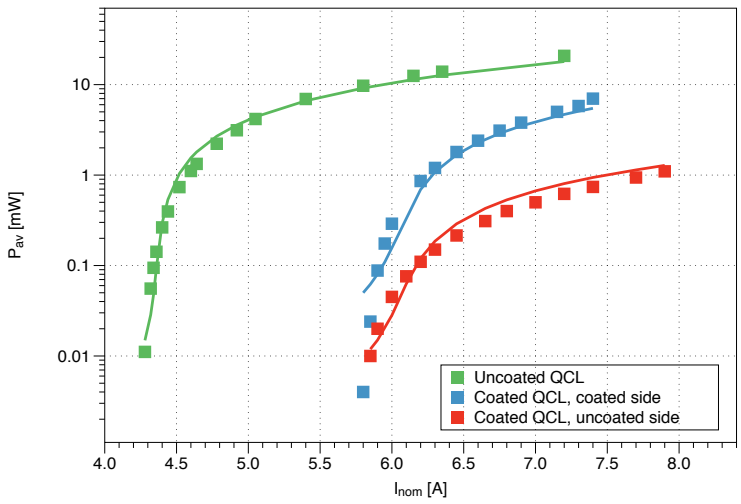


Figure 2.15 *Same P-I curves as in Fig. 2.8 but in a logarithmic plot for better readablitiy. The corresponding lines are the simulated curves using the model.*

Next, we have simulated both P-I curves for the two facets of the coated QCL together, fixing the parameters found for the coated QCL and opening up to the fit procedure only the parameter R_2. The results of this simulation along with the measured values are also plotted in Fig. 2.15 (blue line for the coated facet, red line for the uncoated facet, the dots represent the measurement). It turns out that in order to achieve good results, the saturation photon flux ϕ_s also needs to be opened to the fitting process. This does not seem worrying, because

as the model ignores all effects of spatial hole burning (as discussed in Section 2.5.7), a very different short-range structure of the light may have great influence on the effective value of the saturation photon density. And the nature of the wave inside the resonator looks very different before and after coating. Before the coating is deposited, the resonator is symmetric and a reasonable part of the photons get reflected at either side. Thus, the light strongly resembles a standing wave. After the coating is deposited, the resonator is essentially open and the photons traveling in both directions do not nearly balance each other out, thus the wave is essentially traveling in nature within the entire resonator. The saturation photon flux is fitted to $\phi_s = 27.6 \times 10^{24}$ photons $(\mathrm{cm}^2\ \mathrm{s})^{-1}$ after coating. In order to verify that this is the correct parameter to vary, in a second run, we opened up all of the parameters fitted for the coated laser to the fitting process. It turns out that within narrow tolerances, the other parameters (without ϕ_s and R_2) of the coated laser stably reproduce the values of the uncoated laser. From this we conclude that all the relevant information is preserved in the coated laser, and simulating the uncoated laser and extracting the parameters just serves as a very valuable sanity check for the overall process. This does not contradict the discussion on under-determination of the model, since now *two* curves are fitted at the same time, giving twice as much information, but only one extra parameter R_2 needs fitting.

Finally, the residual reflectivity of the coated facet was found to be $R_2 = 1.54\%$. This value lies within the range of possible values discussed in Section 2.5.3.

2.5.7 Connection with real-world observables

The model presented in the previous sections is a very generic laser amplifier dynamics model that discusses self-consistent, steady-state photon flux densities inside a resonator of a two-level laser. It is useful for fitting effective resonator parameters of most kinds of lasers to experimental data and determining from these some actual facet reflectivities. It can also be used to simulate realistic P-I curves and from this information, the optimal reflectivities for a given application. The only basic assumptions of the model are that, within any range of interest, the population density of the upper laser state is proportional to the population density of the lower state, $N_2 \propto N_1$, and that the pump rate is proportional to the pump current, $\hat{R} \propto I$.

If the parameters that result from fitting experimental P-I curves are to be used not only as phenomenological quantities but to help determine real-world QCL properties, the simple photon propagation model needs to be connected to QCL theory. This can be done through some straight-forward substitutions. For this, it is best to start with the gain.

Eq. (2.51), with Eq. (2.49) and Eq. (2.41) (assuming a normalized Lorentzian lineshape $l(\nu)$ with width $\Delta\nu$), gives a small signal gain of

$$\begin{aligned}\gamma_0(\nu) = N_0\sigma(\nu) &= (\hat{R}_2\tau_2(1-\frac{\tau_1}{\tau_{21}})+\hat{R}_1\tau_1)\frac{\lambda^2}{8\pi t_{\mathrm{sp}}c}l(\nu)\\ &= (\hat{R}_2\tau_2(1-\frac{\tau_1}{\tau_{21}})+\hat{R}_1\tau_1)\frac{\lambda^2}{8\pi t_{\mathrm{sp}}c}\frac{\Delta\nu/(2\pi)}{(\nu-\nu_0)^2+(\Delta\nu/(2\pi))^2}. \quad (2.69)\end{aligned}$$

For QCLs, the small signal gain is given by Eq. (1.29), with Eq. (1.15) and Eq. (1.14), to be

$$\gamma_0(\nu) = J\tau_3(1-\frac{\tau_2}{\tau_{32}})\frac{4\pi e z_{32}^2}{\lambda_0\varepsilon_0 n_{\mathrm{eff}}L_p 2\gamma_{32}}\frac{(\gamma_{32}/2)^2}{(hc)^2(\nu-\nu_0)^2+(\gamma_{32}/2)^2}. \quad (2.70)$$

Connection between the propagation model and the QCL observables means identifying the quantities in Eq. (2.70) with the ones in Eq. (2.69).

First, looking at the assumptions that lead to Eq. (2.45) and Eq. (2.46), it is clear that in a QCL, $\hat{R}_1 = 0$, because the lower laser level is not actively pumped. Also, the levels are renamed from $3 \mapsto 2$ and $2 \mapsto 1$ so that $(\tau_1, \tau_2, \tau_{21}) \mapsto (\tau_2, \tau_3, \tau_{32})$.

The pump rate of the upper level is $\hat{R}_2 = \frac{J}{eL_p}$, because that is the number of carriers injected into the volume of a QCL period per unit time. Then, the small-signal population difference is $N_0 = \frac{J}{eL_p}\tau_3(1-\frac{\tau_2}{\tau_{32}})$. This means that the transition cross section of the propagation model

$$\sigma(\nu) = \frac{\lambda^2}{8\pi t_{\mathrm{sp}}c}\frac{\Delta\nu/(2\pi)}{(\nu-\nu_0)^2+(\Delta\nu/(2\pi))^2} \quad (2.71)$$

corresponds to a transition cross section for QCLs

$$\sigma(\nu) = \frac{4\pi e^2 z_{32}^2}{\lambda_0\varepsilon_0 n_{\mathrm{eff}}2\gamma_{32}}\frac{(\gamma_{32}/2)^2}{(hc)^2(\nu-\nu_0)^2+(\gamma_{32}/2)^2}. \quad (2.72)$$

But this is verified easily with some simple considerations. Both gains were assumed to be homogeneously broadened, so for simplicity, only the peak values are compared, which are

$$\sigma(\nu_0) = \frac{\lambda^2}{8\pi t_{\mathrm{sp}}c}\frac{2}{\Delta\nu\pi} \quad (2.73)$$

and

$$\sigma(\nu_0) = \frac{4\pi e^2 z_{32}^2}{\lambda_0\varepsilon_0 n_{\mathrm{eff}}2\gamma_{32}}. \quad (2.74)$$

First, $1/t_{\mathrm{sp}} = A_{21}$, where A_{21} is the Einstein A Coefficient of the transition. But $A_{21} = B_{21}\frac{8\pi h n^3}{\lambda_0^3}$, where B_{21} is the Einstein B Coefficient. Also, $B_{21} \propto \frac{|M_{21}|^2}{\varepsilon_0 h^2}$, where M_{21} is the transition Matrix element (the unitless proportionality constant depends on the exact nature of the E-field). But obviously, $|M_{21}|^2 = (ez_{32})^2$. Also, the width of the Lorentzian in Eq. (2.72), γ_{32}, is in units of energy and the width in Eq. (2.71) is in units of wavenumber, so the substitution $\Delta\nu = \frac{\gamma_{32}}{hc}$ has to be made.

With the above, it is also clear that the gain coefficient $\bar{g}$ from the propagation model in Eq. (2.65) corresponds to the gain coefficient $g(\nu)$ for QCLs in Eq. (1.16) and Eq. (1.29),

where one is defined as the proportionality factor between pump current and gain and the other is pump current *density* and gain.

Finally, one last correction has to be made to the gain with regards to the overlap factor Γ. Γ was introduced in Eq. (1.20) as a phenomenological factor that increases the losses, since it is assumed that the part of the light field that does not overlap with the active region is lost entirely. However, in the photon propagation model, the mirror losses are not entered as factors that clamp the gain, but they appear dynamically through the boundary conditions and the self-consistency requirement. In fact, the correct waveguide loss α_w is best entered manually, but then the mirror losses (through the facet reflectivities) are stably fitted to the correct values. However, it is clear that the gain value that balances the losses is then an effective gain value because it, so far, does not consider the overlap factor. This can be corrected by making the transition $\bar{g} \mapsto \bar{g}\Gamma$ (obviously, Γ needs to be known from external measurements). This is equivalent to the statement that only the light in the active region experiences gain. But then, finally the gain coefficient from Eq. (2.65) $\bar{g} \mapsto \frac{g}{lb\Gamma}$ from Eq. (1.29).

Although the model expands on the standard treatments by allowing for a spatial variation of the gain saturation and includes phenomenological spontaneous emission, it still has strict limitations as mentioned earlier. Most importantly, it discusses slowly-varying, transversely homogenous photon flux densities inside a resonator and thereby completely ignores the wave-like nature of the electric field. This has, among others, the following consequences. The light does not form standing waves, thus the model is oblivious to modes or any other effects of phase-dependence. It is consequently an inherently single-mode model. There are also no nodes and there is no short-range longitudinal structure, thus no longitudinal spatial hole burning. There is also no transverse modal structure or transverse hole burning and there is no mode competition. Since the model is, in its simplest form, a single-wavelength model, there are no homogeneous or inhomogeneous broadening effects. It is a time-independent, steady-state model, thus it covers no transient effects, thus it is in principle not applicable for short pulses. It is also oblivious to all temperature effects. This includes the effects of the heatsink temperature, as well as heating of the active region. And finally, it ignores all microscopic QCL-specific details such as multiple laser levels with different broadenings, thermal leakage and backfilling, etc.

These limitations can however be selectively amended to any level of desired physical accuracy. Either by phenomenologically adding an arbitrary number of modes that get treated in parallel within the rate equations and the self-consistency checks, or by adding more physical detail, such as temperature, time, or phase-dependence. This however will result in a much greater demand for computing power, and is unnecessary for our purposes.

2.5.8 Some remarks about the gain

The condition that the gain is a constant throughout the entire resonator is equivalent to the statement that the reflectivities of the end mirrors are nearly 1. Since if no light is lost at each end but is reflected in total, the steady-state photon density distribution inside the resonator is just a constant. But with gain and loss present, this can only be the case if the (saturated) gain exactly equals the distributed loss, and thus, it is also a constant. Relation (1.7) can also be understood as an *effective* round-trip gain even in case the reflectivities of the facets are low. However, the canonical form contains no information about the light emitted below the threshold, and therefore is of no use when characterizing the light emitted by a QCL with a good AR coating.

However, even if the reflectivities are approximately 1, the gain is only a constant for each mode individually, but not for all modes at the same time. This is clear from the following. Due to the strongly pronounced nodes of the standing waves, there are alternating regions of high and low photon flux density. But gain saturation just states that gain is smaller in ranges where there is more light. Thus the gain of the resonator has an undulating structure with the same period as the standing wave, with much higher values at the nodes of the standing wave, which consequently cannot be "seen" by that laser mode. However, laser modes that are a sufficient number of longitudinal orders higher or lower, thus have nodes spatially removed, can experience gain without competing with the first mode. This effect is called (longitudinal) spatial hole burning, because the first mode (reversibly) "burns" holes into the gain medium, i.e. selectively saturates the gain. Similarly, higher transverse modes can also exist, since their peak photon densities are also spatially removed from the peak densities of the fundamental mode.

Due to the oscillation of the light density at the apexes of the standing wave, it is worth considering, whether there is a time dependence to the gain. But it turns out that the oscillation period is on the order of 10 femtoseconds, while even for an ultrafast QCL, refilling of any laser state is on the order of picoseconds and thus smoothes out any possible time dependence.

Especially if the reflectivities are not close to 1 and the gain is spectrally broad enough to allow for multiple modes, the gain is a very complicated function. It contains a slowly varying envelope function (considered in the photon flux model) that is a constant only in the limit of high reflectivities and is different for each longitudinal and transverse mode due to hole burning. If the reflectivities are smaller than 1 and equal, the gain has a maximum in the center of the resonator. The more pronounced the nodes of the standing wave, the more pronounced the short-range spatial oscillation of the gain. If the reflectivities are low, the light distribution has the shape of a traveling wave towards the ends of the resonator and gradually forms into a standing wave in the center of the resonator.

If there is more than a single upper and/or lower laser level that are filled independently, the above possibilities are multiplied with the number of possible laser transitions.

Thus finally, it is clear that a simulation of the full laser dynamics with given laser transi-

tions is computationally very intensive and simplifications are often highly advisable.

2.6 In-Situ Evaluation Procedure

As mentioned in Section 2.4, the deposition system for AR coatings contains a (quasi) in-situ measurement setup for the P-I characteristic of the laser. In that sense, the analysis of the optical coating is performed while the coating is deposited, and the setup of the setup of the coating process already contains the expected behavior of the finished result.

The in-situ evaluation of the layer thickness, i.e. coating quality during deposition is best done in two stages. In the first stage, the residual reflectivity of the deposited layer can easily be determined from the rising threshold current of the QCL, as long as there is a clearly defined threshold current. This works as follows. Eq. (2.32) states that the threshold current density of a QCL is

$$J_{\mathrm{th}} = \frac{\alpha_w + \alpha_{m1} + \alpha_{m2}}{\Gamma g}, \tag{2.75}$$

where $\alpha_{mi} = \frac{1}{2l} \ln \frac{1}{R_i}$ are the mirror losses of facet i, Γ is the confinement factor, i.e. the overlap of the laser mode with the active region, and g is the gain coefficient defined by $gJ = \gamma$. Inserting the mirror losses and cross sectional area of the current flow, $l \times b$, where b is the width of the laser stripe, Eq. (2.75) reads

$$I_{th} = lb \cdot J_{th} = lb \cdot \frac{\alpha_w + \alpha_m}{g\Gamma} = \frac{lb}{g\Gamma}(\alpha_w - \frac{\ln R_1}{2l} - \frac{\ln R_2}{2l}). \tag{2.76}$$

If α_w is known, $g\Gamma$ can be extracted from a P-I curve measured for the uncoated laser. R_2 of Eq. (2.76) is now to be understood as the residual reflectivity of the coated facet.

Eq. 2.23, giving the reflectivity of a layer of some material of thickness d and refractive index n deposited on a substrate with refractive index n_s for a given wavelength λ at normal incidence in ambient vacuum or air, can be rewritten to

$$R = 1 - \frac{4}{2 + \frac{1}{n_s} + \frac{n_s}{1} + (n_s - \frac{n^2}{n_s})(\frac{1}{n^2} - 1) \sin^2 \frac{2\pi n d}{\lambda}}. \tag{2.77}$$

For the case of $n^2 = n_s$ and $d = \frac{\lambda}{4n}$, R reduces to 0. This is the ideal case. If $n^2 \approx n_s$, then $R \approx 0$ but finite at the same thickness d. Next, Eq. (2.77) is substituted for R_2 of Eq. (2.76) to give the threshold current as a function of layer thickness d. This is plotted as the resonance curve in Fig. 2.16, together with actual measurement results during a coating run marked as dots.

The second term on the right-hand side of Eq. 2.77 has the shape of a resonance curve in d. This plus the fact that R_2 of Eq. (2.76) is log-transformed, lead to the fact that this measurement is very sensitive to R_2 as R_2 approaches 0.

However, as discussed in Section 2.5.3, at some point during deposition, the picture of a clearly defined threshold current loses validity. But instead of seizing emission entirely, as would be expected for an ideal laser below threshold, a considerable amount of emission

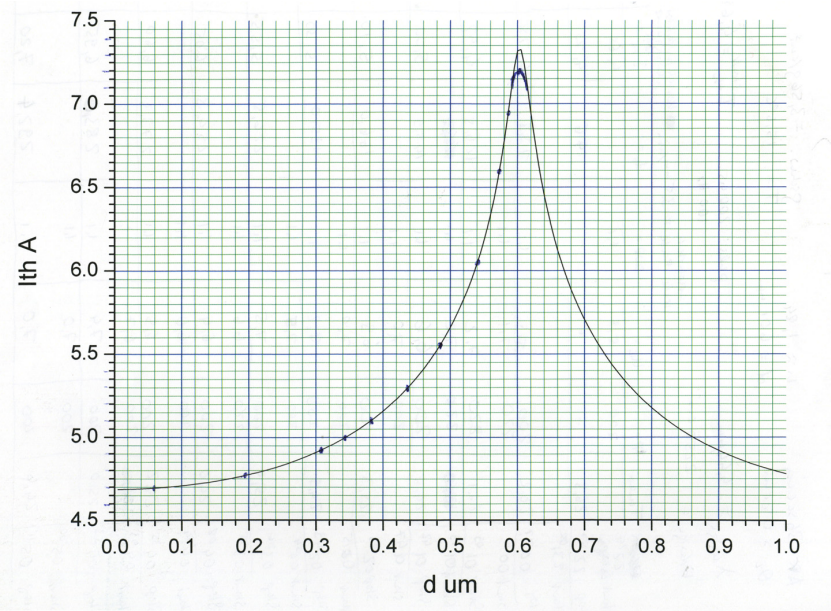


Figure 2.16 *Protocol of the threshold current measurements during the deposition of a quarter-wave Y_2O_3 AR coating on a QCL facet. The curve is calculated with Eqs. (2.76) and (2.77) with a design wavelength of 4.44 µm and index of 1.843. The dots mark the measured values of the threshold current after intervals of deposition.*

remains. The response just misses the characteristic "kink". This is because, as also discussed in previous sections, the former laser turns into a superluminescent diode (SLD), which, due to the long stripes of often around 6mm, causes considerable emitted powers. This makes interpretation of the in-situ results rather hard, and the best way around it is the following.

It is clear from Eq. (2.39) and Fig. 2.8, that the "slope" of the uncoated facet's P-I curve, even if the picture of a laser is not strictly valid, decreases with decreasing reflectivity of the other facet. Consequently, the best way to reach the lowest possible reflection is to follow carefully the decrease of emitted power at the *highest allowable pump current value* during deposition. Since this does not translate in any simple way to the residual reflectivity, the only way to find the optimal value is to measure in gradually decreasing intervals and to observe when the emission minimum is *passed*. The final reflectivity can then only be determined *ex-situ* by perfoming P-I measurements on both facets and fitting the results using the photon flux density model.

The first stage of the in-situ evaluation can also be performed when depositing two-layer coatings. Instead of Eq. (2.77) it is necessary to use Eqs. (2.25), (2.26), and (2.27) from Section 2.3. The challenge with this is that the threshold current is very insensitive to changes of the overall reflectivity R_2 as long as R_2 is not very small. Therefore, the actual layer thickness of the first layer is strongly prone to error. Extracting the thickness from the reading on the quartz oscillator is also completely futile, since the tiny facets are directly on the edge of the substrate and edge effects dominate the deposition of the coating material onto it. These in turn dependent strongly on exact facet size, whether the structure is overgrown, and the thicknesses of the insulating layer and gold overcoat. From this it follows, that the

best way to make two-layer anti-reflection coatings is by using two materials that are both as close as possible from two sides to the ideal one-layer value. The more this is the case, the more tolerant the process is to thickness errors of the first layer, since the error of the first layer can be balanced out by the second. The final stage of the deposition can then also be monitored by the maximum-power method, which once overall reflectivity is small becomes very sensitive.

3

Versatile Littrow External Cavity QCL with Reflective optics (V-ECQCL)

3.1 Introduction

The first External Cavity (EC) CQL developed for this work is a Littrow cavity that employs reflective optics to collimate the beam from the AR-coated intra-cavity QCL facet. The collimating optical element in this case is a comparatively large 1"-diameter, gold-coated, 90° off-axis parabolic mirror with an effective focal length of 1", mounted on a 6-axis kinematic mount.

A great advantage of using purely reflective intra-cavity optics (mirrors) in an EC, is that sufficiently accurate alignment can be performed using a visible pilot beam coupled into the optical path, regardless of the actual *wavelength* emitted by the semiconductor laser. Refractive optics (lenses) often prohibit this, as the materials employed in one wavelength range are normally not transparent in another. See for example silicon, germanium, and the chalcogenide glasses, which are normally used in the MIR, but are opaque in the visible range (VIS). And if a material is transparent in both the VIS and the laser's emission range, it would most certainly have a different index of refraction in the two, making a lens quite difficult to align correctly using the visible beam, since the focal point would be shifted. A parabolic mirror, however, has a focal point that is oblivious to the wavelength in use, as long as it is one for which the metal coating is reflective. Thus, depending on the metal of the coating, e.g. gold, silver, or aluminum, the same setup can be used for semiconductor lasers emitting anywhere between the terahertz (THz) and ultra-violet (UV) ranges.

An alignment procedure using a visible beam, however, has another great advantage, as it can be used to collimate a laser's output regardless of its *power*. But this is not only a convenience, but a requirement for setting up Littrow ECs with low power lasers. The reason for this is as follows. Applying the necessary AR coating to one of the laser facets (to eliminate

mode-hopping phenomena) reduces the power emitted by the laser substantially if there is no external feedback. In the case of very good AR coatings, it reduces the output to (amplified) spontaneous emission only. Now, in ranges where high-power lasers are available, collimation is straight forward. Because if the power is high enough for there to be a reasonable amount of light left even without external feedback, it can be used to align a collimating lens with the use of a power meter, a camera, IR viewing cards, or even the naked eye in the case of visible light. But once the beam is collimated, alignment of the grating is easy, as is discussed shortly. However, the availability of high-power semiconductor light sources is limited to more or less small patches of the total range spanning from the THz to the UV. Everywhere else, where the sources are more experimental and the achieved powers low, the emission is often not sufficient to align a collimating lens accurately in the conventional way.

Especially the THz range poses added difficulties to being aligned in the conventional way. Due to the large wavelength of the THz, the diameter of the collimated beam needs to be very large in order for the feedback from the Littrow grating to have a small bandwidth. This is because as the wavelength scales, the groove distance of the Littrow grating scales along proportionally, because for the Littrow principle to work, the grating is employed in the first interference order. But the bandwidth of a grating in the Littrow configuration scales with the number of illuminated grooves. Thus the necessary beam diameter for a given grating selectivity also scales with wavelength. Thus, in order to achieve the same selectivity in the THz that a lens with a 2-3 mm aperture has in the MIR, beam diameters of at least 1" are necessary. But the facet dimensions of THz QCLs, especially parallel to the growth axis, are only a few microns, thus compared with the focal distance of the optic, they are minuscule, making the slightest angular alignment error catastrophic. At least one facet dimension is also smaller than the wavelength, but Gaussian optics forbids arbitrarily small focal points for light of finite wavelength. Thus even at perfect alignment, only a small fraction of light reflected from the grating actually couples back into the facet. So since even a perfectly aligned THz EC has little feedback, it is ever harder to align, and it is more important to align it very well. But THz QCLs need to be cooled to very low temperatures to operate, so the EC would need to be aligned inside a cryostat. Which seems unrealistic given these challenges. Consequently, to our knowledge, conventional Littrow EC QCLs in the THz have not been demonstrated so far, however there are a number of more or less elaborate workarounds in the literature [134].

As mentioned earlier, the difficult part of aligning ECs with low-power laser sources, or simply sources that are switched off, such as THz lasers outside a cryostat, is to achieve good collimation, since once the beam is well collimated, achieving feedback from a grating using the emitted beam is generally quite simple. This is because for the grating, only one tilt degree of freedom needs to be varied. The other, the tuning angle, can normally be simply set by hand without fine adjustment, because the tolerance on this angle is large, since as long as *any* wavelength of the gain fulfills the Littrow condition, this angle is sufficiently adjusted. Thus, to align the grating, one angle is set coarsely, then the laser is switched on and the last remaining degree of freedom is scanned through until feedback is achieved. But then it becomes quite obvious when feedback is achieved, since at this moment the emitted power

rises by 2-3 orders of magnitude.

The EC setup and its alignment procedure described in the next sections can be used for any semiconductor laser regardless of the wavelength and power of the emission, since it allows reflective collimating optics of arbitrary size to be aligned quickly and optimally with the use of a visible beam only. This makes the EC very useful as a research device, i.e. for testing spectral characteristics of new lasers, AR coatings, and prototyping and investigating tunable lasers with the new sources. This has lead to patent no. DE102012000038A1 [1]. For further reference, this setup will be called the V-ECQCL.

3.2 Experimental Setup and Alignment

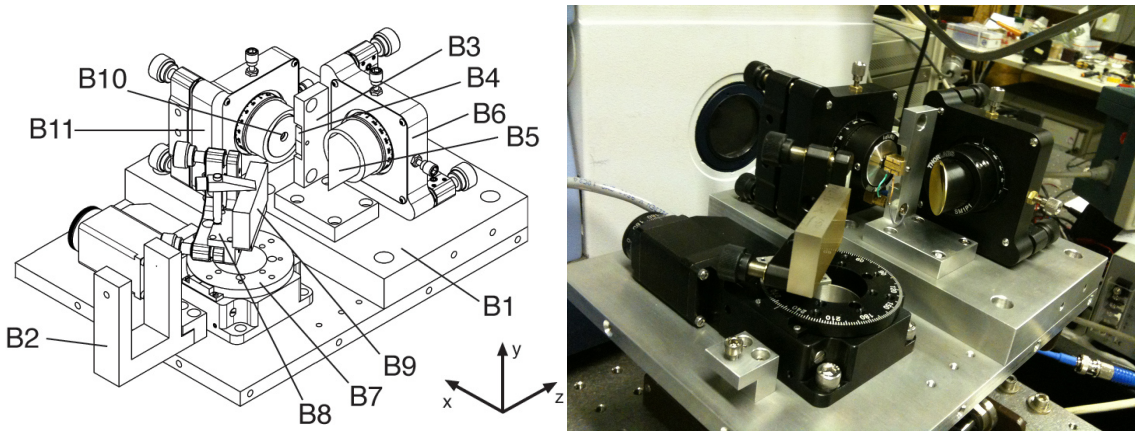
This section shows the experimental setup of the Versatile Littrow External Cavity (V-ECQCL).

The challenges in aligning an External Cavity QCL can be summarized as follows: 1. Small facet; 2. Great beam divergence; 3. Invisible (MIR or other) low-power beam. Furthermore, contrary to popular practice, it is not always advantageous to place collimating intra-cavity collimating optics close to the facet, because great beam diameters are required in order to obtain small line widths, especially at greater wavelengths.

As mentioned earlier, our approach is to couple a visible (diode laser) beam into the QCL's beam path with sufficient accuracy to make a reliable QCL alignment possible with the sole use of optical tools and only one mechanical degree of freedom left to adjust using intensity measurements. To further facilitate this process the mounts of the optical components with their respective degrees of freedom are chosen so as to make a step-by-step alignment possible without extensive iteration cycles. The arrangement consists of an optical bench, photographed in Fig. 3.2, to condition the visible beam, the EC-QCL setup, sketched in Fig. 3.1 a) and photographed in Fig. 3.1 b), and a locking mechanism to define a joint optical axis for the EC-QCL apparatus and the optical bench.

The step-by-step assembly and wavelength-independent alignment procedure of the setup are covered in Appendix D.

The EC-QCL setup B is sketched in Fig. 3.1 a) and photographed in Fig. 3.1 b) and the parts are described in Table 3.1. The optical bench A is photographed in Fig. 3.2 a) and the parts are described in Table 3.2.



(a) *Sketch of the V-ECQCL with annotated parts.* **(b)** *Photograph of the V-ECQCL as seen in a), though without the alidade.*

Figure 3.1 *Setup of the V-ECQCL.*

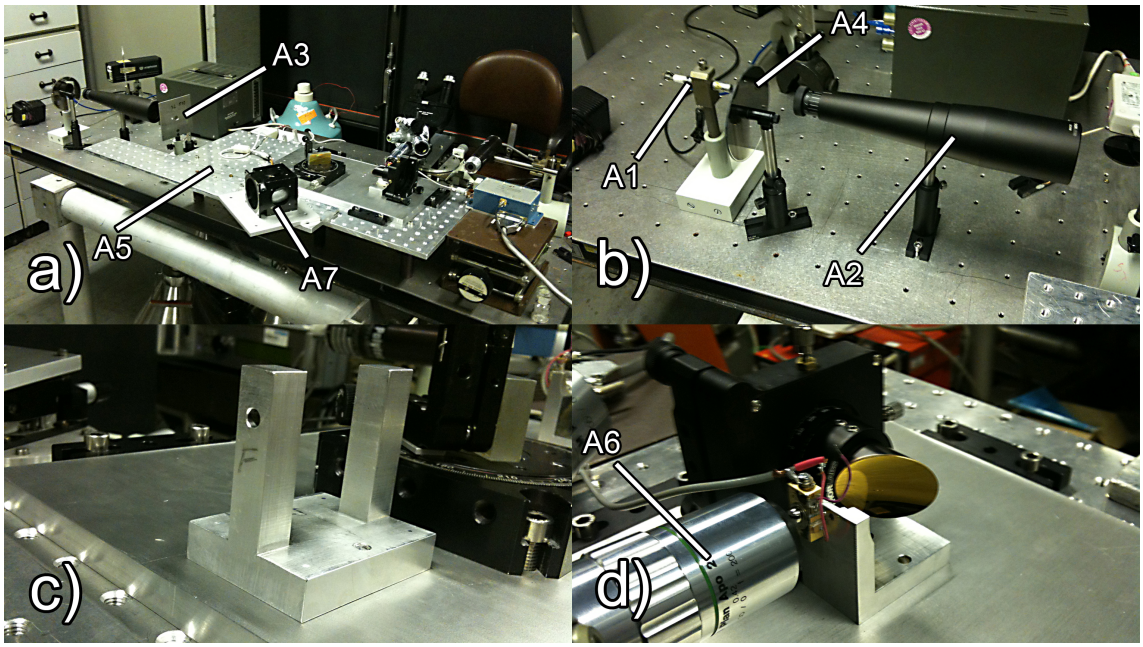


Figure 3.2 *a) The optical bench (with mounted EC setup on the right). b) Close-up of the visible diode laser, grey filter, beam expander (from left to right). c) Close-up of the alidade that consists of two aluminum pillars with 1/10-mm pin holes in each (not visible, only the 5 mm drill hole used to thin down the material is seen on one of the pillars). d) Collimating off-axis parabolic mirror, mounted QCL, and ultra-long distance microscope. The QCL is mounted coaxially with the microscope's objective and its c-mount facing left. The expanded visible beam enters from the bottom-right of the picture, is diverted 90° and focussed on the QCL's intra-cavity facet that faces towards the mirror. The focus point of the visible beam is then viewed with the long-distance microscope's focal plane set to contain the QCL intra-cavity facet.*

Table 3.1 *EC-QCL setup (B)*

B1.	Base plate
B2.	Alidade (double pin hole), detachable, locked to base plate in well-defined position with 0 degrees of freedom, defines an optical axis to within <4 mrad
B3.	QCL mount, fixed, 0 degrees of freedom
B4.	QCL
B5.	Intra-cavity collimating off-axis parabolic mirror, gold-coated, 1" diameter
B6.	6-axis mount for mirror, lockable
B7.	Electric rotation stage for grating (with driving electronics)
B8.	Kinematic mount, high stability, with one high-precision differential actuator (one degree of freedom)
B9.	ML Grating, 1" × 2", gold coated
B10.	Extra-cavity collimating aspheric lens, AR-coated for given wavelength region on
B11.	6-axis mount, lockable

Table 3.2 *Optical bench (A)*

A1.	Visible diode laser (with driving electronics, Fig. 3.2 b) left)
A2.	Beam expander (Fig. 3.2 b) center)
A3.	Circular 1" aperture Fig. 3.2 a) (just to the right of the beam expander)
A4.	Variable gray filter (Fig. 3.2 b) left)
A5.	Guiding rail with locking nuts to securely mount EC-QCL setup (Fig. 3.2 a) right)
A6.	Extra-long working distance microscope, horizontally mounted (Fig. 3.2 d) left and Fig. 3.2 a) right))
A7.	Light for microscope (Fig. 3.2 a) center, inside the black cube)
A8.	Laser power meter (with driving electronics, not pictured)
A9.	Pyrocam (optional, not pictured)

3.3 Housing and Cryostat

The V-ECQCL, when dismounted from the optical bench has a relatively compact footprint of 20 cm×25 cm. To allow the setup to be purged with a neutral gas, e.g. when performing gas spectroscopy experiments, a housing can be attached, as photographed in Fig. 3.3. This also improves the portability of the setup.

For experiments that require cooling, a large-volume cryostat has been designed and ordered. The vacuum chamber of the cryostat fits the entire V-ECQCL and can be cooled using liquid nitrogen. Since the fully aligned EC can be placed inside the cryostat, a number of

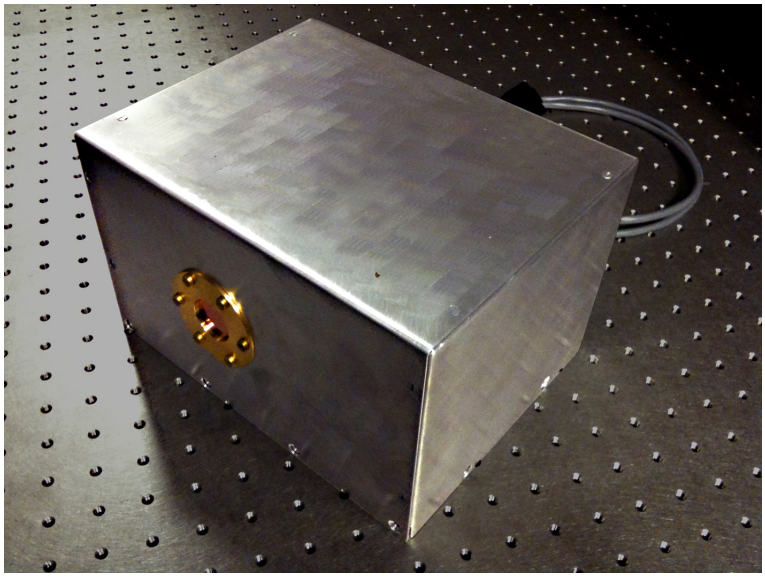


Figure 3.3 *V-ECQCL inside a 20 cm×25 cm×15 cm housing that allows purging the setup with a neutral gas and improves portability.*

different types of experiments can be performed using this setup. First, any experiment that requires cooling the QCL, e.g. measurements with THz QCLs or temperature-dependent measurements of the QCL in the context of an EC can be performed. Second, if a small gas cell is added to the vacuum chamber, temperature-controlled gas spectroscopy measurements can be made. Third, intra-cavity gas absorption spectroscopy can be performed due to the fact that the ambient air of the EC can be replaced by a temperature, pressure, and composition-controlled mixture of gases.

This rather unique setup has proven to be a very useful research tool for a number of different experiments not part of this thesis.

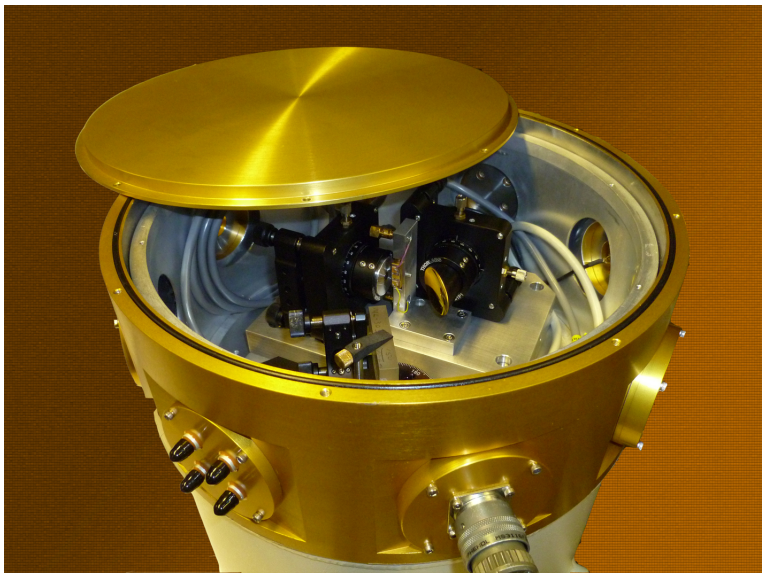


Figure 3.4 *Large-volume cryostat. The vacuum chamber of the cryostat fits the entire V-ECQCL and can be cooled using liquid nitrogen.*

3.4 V-ECQCL in conjunction with photoacoustic detection used as a gas sensor

In this section we describe a simple and compact gas sensor, that is composed of the V-ECQCL operating at room temperature and a detector. We compare the performances of detection with a MEMS-scale miniature resonant photoacoustic cell and, alternatively, absorption within the free space of the external cavity. For the two different systems we use two QCLs emitting in different regions. These configurations were used for ammonia detection near 11.3 μm and CO_2 isotopologue recognition in human breath near 4.35 μm. Mode discrimination in the external cavity is discussed in conjunction with the transmissivity of the antireflection coating of the laser facet and the geometric design parameters of the external cavity setup.

The sensors neither employ temperature control of the QCLs to achieve mode-hop free behavior, nor piezo-controlled cavity lengths for continuous tuning. Instead, an approach of optimizing the AR coating and the geometrical parameters of the EC setup is chosen. The mode-hop related intensity fluctuations can be suppressed through sufficiently low tolerance antireflection coatings and the resolution is increased through an unusually large intra-cavity beam diameter to maximize the grating selectivity. This results in resolutions better than 0.2 cm^{-1} and sufficient detectivity to resolve the mixing ratios of ^{12}C and ^{13}C isotopologues of ambient CO_2. The footprint of the sensor (without electronics) is 20 cm x 25 cm.

The QCLs have been implemented with and without antireflection coatings to compare performances.

3.4.1 Experimental

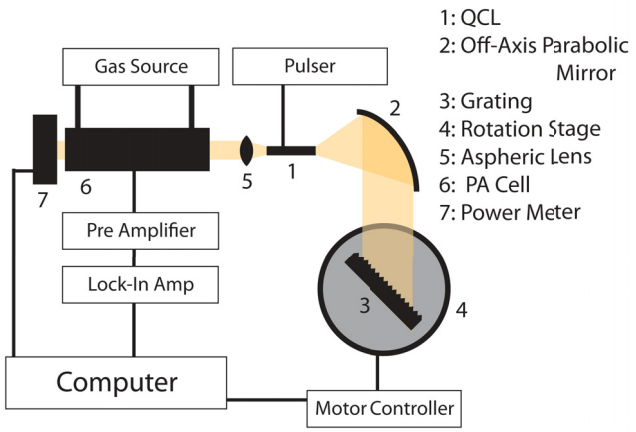


Figure 3.5 *Experimental setup.*

The experimental setup is illustrated in Fig. 3.5. The Littrow-type external cavity setup consists of the QCL, a collimating 90° off-axis parabolic mirror (Edmund Optics) with an effective focal length and diameter of 1″ (2.54 cm). The feedback was provided by a 1″ x

2″ optical grating (Optometrics) with groove spacings and blaze wavelengths appropriate for the different wavelengths. The grating was mounted on a motorized rotation stage (Standa, 8MR151) with a resolution of 4.5 arcsec/step. To maximize output power, the useful light was coupled out of the uncoated facet facing away from the grating. The as-cleaved facets have a Fresnel-reflectivity of approximately 28%. The output beam was collimated with an aspheric chalcogenide lens of 5.5 mm diameter (Edmund Optics). The QCL was powered by an Avtech pulser and the pulse shape was monitored with an oscilloscope. The current was chosen so that the pulse peak amplitude lies just below the threshold for self-oscillation on the FP modes to ensure no lasing takes place outside the grating-selected modes. There was no active cooling or active mechanical stabilization. For power measurements we used a bolometer with a lock-in frequency of 30 Hz and an integration time of 1 s. The details of the photoacoustic cell can be found in Ref. [135].

3.4.1.1 CO_2 Detection at 4.3 μm

CO_2 detection was based on this EC configuration using a 40-cascade QCL chip emitting at 4.35 μm at room temperature with the active region design (our ref.: HU2-0324) based on the composite-barrier approach (Ref. [11]) with the band alignment illustrated on Fig. 3.6. The major improvement of the current design over the one in Ref. [11] is the increased (up to 80 meV) energy spacing between the upper laser level (upper black line on Fig. 3.6) and the bottom of the miniband of excited states (thin grey lines on Fig. 3.6).

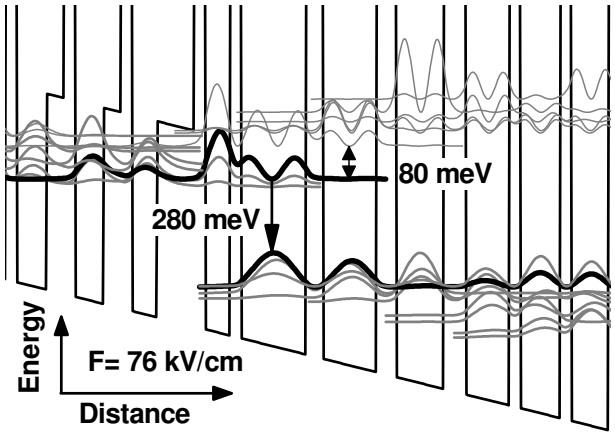


Figure 3.6 *Conduction band diagram and probability functions calculated within a single period of the active region with a 76 kV/cm electric field. The layer thicknesses in nm from left to right starting from the injection barrier (thickest composite barrier) are: 3.0/**0.9**/2.0/**0.9**/5.2/**1.4**/4.3/**1.6**/3.8/**1.8** /3.5/**0.9**/3.2/**0.9**/3.0/**0.9**/2.4/ 1.4/**0.9**/2.3/1.4/**0.9**/2.0. AlAs layers are in bold, $In_{0.73}Ga_{0.27}As$ layers are in roman, and $In_{0.55}Al_{0.45}As$ layers are in italics. Underlined layers are doped to 5×10^{17} cm^{-3}. The moduli square of the wavefunctions responsible for the laser transition are drawn with the thick black lines. Ground injector/extractor states are drawn with the thick gray lines and the lowest excites states are drawn with the thin gray lines.*

We coated one facet of this QCL with a 605-nm single-layer AR-coating of Y_2O_3 that we have measured to be transparent at this wavelength with a refractive index of 1.84 [192]. The coating has a measured residual reflectivity of less than 1.5% as we determined in Section

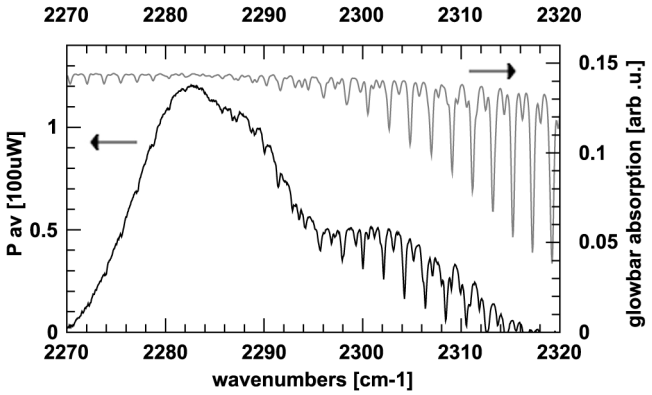


Figure 3.7 *Output power of the 4.3 μm Laser as a function of the grating-selected wavenumber (center of bandwidth) for a drive current of 7.2 A plotted along with a glowbar spectrum of the same spectral range to clarify that the fluctuations in power are not due to FP modes but ambient CO_2 absorption lines instead. The lines to the left of 2284 cm^{-1} are ^{13}C-isotopologue CO_2 lines.*

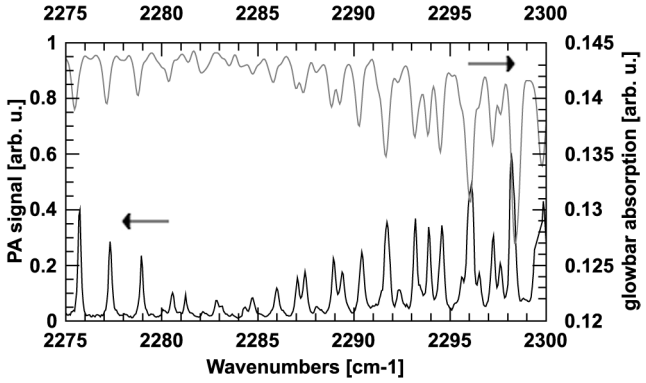


Figure 3.8 *Corresponding PA signal to Fig. 3.7 (plotted for reduced range) when used in conjunction with a PA cell to detect CO_2 in human breath.*

2.5.6. Feedback in the external cavity was provided by a grating with 300 grooves/mm (ML 601) with a blaze wavelength of 3.5 μm and nominal efficiency of >80%. Figure 3.7 shows the output power as a function of the grating-selected wavenumber (center of bandwidth) for a drive current of 7.2 A with a pulse width of 100 ns and a repetition rate of 80 kHz. The tuning range is 48 cm^{-1} and the maximum power is 120 μW. These values can be further increased by optimizing the active region design.

Actively controlling the QCL temperature can be used to compensate for intensity fluctuations related to mode-hopping in between Fabry-Perot (FP) modes of the chip. However, this slows down the scanning speed. Figure 3.7 shows that at least for this wavelength region it is possible and straight-forward to produce an AR coating with sufficiently low residual reflectivity to render these fluctuations negligible. The dips in power in the data of Fig. 3.7, are instead due to intra-cavity absorption by ambient carbon dioxide in the lab air over the free-space path length of $\approx$ 9 cm. This is clarified by plotting an FTIR transmission spectrum of the same lab air with a glowbar as the broadband light source in the same graph to facilitate comparison of peak positions and shapes. The dips at wavenumbers less than 2284 cm^{-1} actually correspond to the ^{13}C isotopologue of CO_2 whose concentration is approximately 4 ppm in ambient air. The spectrum was taken with an integration time of 1 s per point.

This shows that even without use of a photoacoustic cell, the EC setup functions as a relatively sensitive detector, with its detectivity scaling exponentially with cavity length. This sensitivity can be further increased by coating the output facet to have a higher reflectivity to increase the finesse of the cavity, meanwhile optimizing the active region to produce higher powers to be measured in spite of the increased reflectivity of the output facet.

Finally, the setup was used in conjunction with a photoacoustic cell, this time to measure carbon dioxide in exhaled breath with a concentration of approximately 4% at atmospheric pressure. The results of this measurement are in Fig. 3.8. Spectral resolution of better than 0.2 cm^{-1} is achieved here. As discussed in section (3.4.2), the spectral resolution can be increased by decreasing the drive current, since only one longitudinal mode of the EC is driven above threshold at the expense of lowering the output power.

In the current configuration, we use a beam impinging on the PA cell with a diameter of 5 mm. Using a beam with diameter of 1 mm will allow us to use a successor to the current PA cell (not published) with a sensitivity of two orders of magnitude greater.

3.4.1.2 Ammonia Detection at 11.1 μm

For the absorption spectroscopy of ammonia, we used a 25 μm wide, 100-cascade QCL, with a facet height of 5 μm and a length of 5 mm, emitting in the range around 11.1 μm (900 cm^{-1}). Details of this laser can be found in Ref. [136].

We used a double-layer AR coating composed of 859 nm of Y_2O_3 with a measured refractive index of $1.47 + i0.02$ [192] and 142 nm of Si with a refractive index of 3.42 deposited with rf-sputtering. A cross-sectional SEM can be found in Fig. 5.20. The power as a function of drive current (P-I) for this chip for both the coated and uncoated facets after deposition are plotted in Fig. 3.10. From these curves we determined the residual reflectivity of the AR

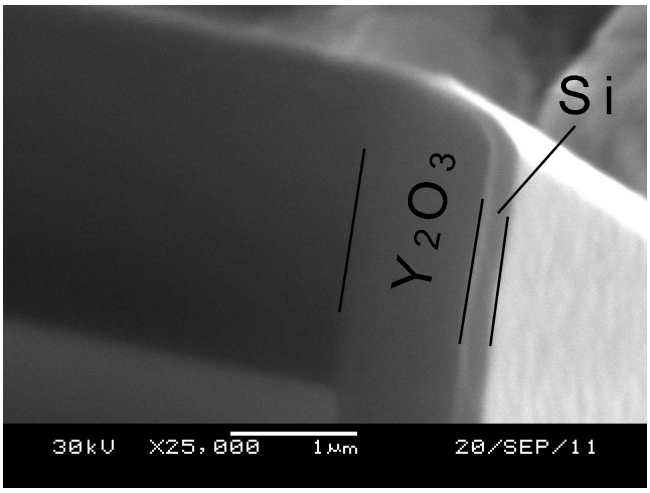


Figure 3.9 *Crossectional SEM of the AR coating for 11.1 µm. Clearly visible is the thick Y_2O_3 layer and the thinner Si layer on top.*

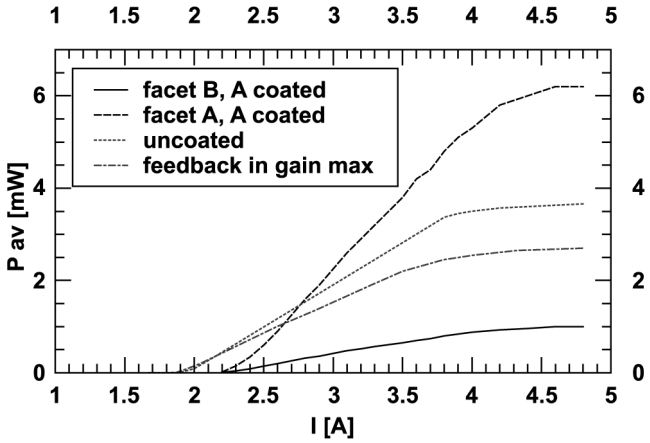


Figure 3.10 *P-I curves for the QCL emitting at 11.1 µm (2) taken with a pulse width of 100 ns at a duty cycle of 0.2%. Plotted are the curves for the uncoated laser, both facet powers for the AR coated laser, and the power output when run in the EC with the grating tuned to the maximum of the gain region.*

coated facet to be approximately 9.0%.

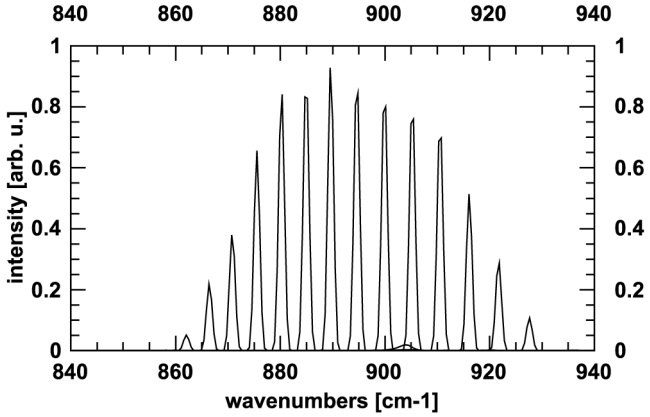


Figure 3.11 *Selection of spectra, taken at different grating angles. Resolution limited by the FTIR.*

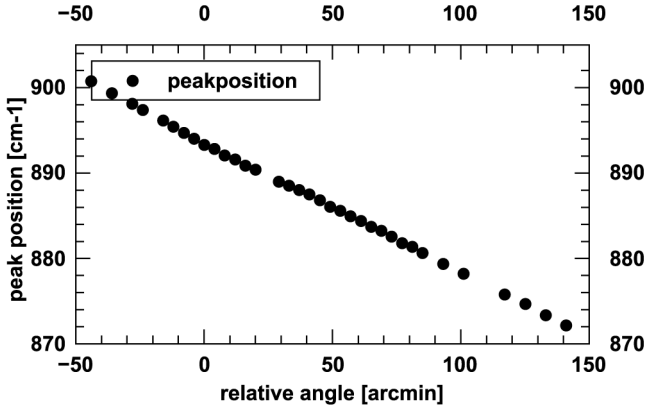


Figure 3.12 *Selection of peak positions as a function of their corresponding grating angles.*

This active region was incorporated in the external cavity, feedback was provided by a 150 grooves/mm grating (ML 303) with a blaze wavelength of 10.6 μm and nominal efficiency of >88%. Figure 3.11 shows a selection of spectra, taken at different grating angles; these spectra are limited by the resolution of the FTIR (Bruker). A selection of peak positions as a function of their corresponding grating angles is displayed in Fig. 3.12. The highest average power of 2 mW on the grating-selected mode in the center of the tuning range was reached with a pulse width of 100 ns and a repetition rate of 200 kHz.

This setup was used in conjunction with the photoacoustic cell to measure the absorption spectra of ammonia diluted in nitrogen with a concentration of 0.015 at atmospheric pressure. The results of this measurement in comparison with the appropriate HITRAN [137] values can be seen in Fig. 3.13. A linewidth of 0.2 cm^{-1} is measured, which is the pressure-broadened ammonia absorption line width at atmospheric pressure. The detectivity of this setup is limited 800 ppb due to the low power resulting from the imperfect AR coating. For the same reason, power fluctuations are clearly visible when comparing the measurement with the HITRAN curve, but the overall shape of the absorption lines are in good agreement. For an analysis of the power fluctuations, see the next section. The beam impinging on the PA

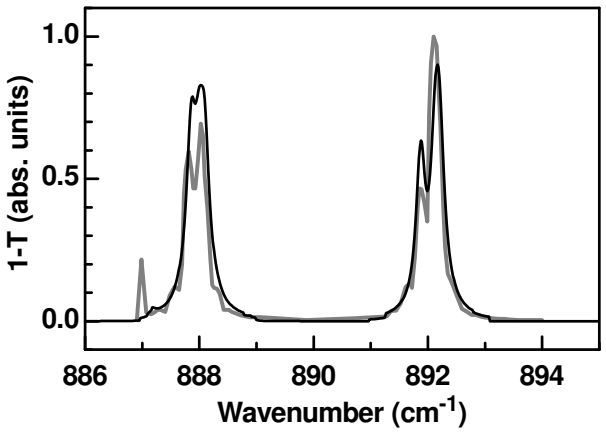


Figure 3.13 *(Black line) Absorption spectrum within the length of the photo-acoustic cell, calculated using 0.015 value mixing ratio at atmospheric pressure, using the HITRAN database (Ref. [137]). (Gray line) scaled response signal of the photo-acoustic cell excited with EC QCL in the given spectral range of symmetric bend vibration. Power fluctuations (see next section) are clearly visible in the measured curve, but the overall shape of the absorption lines are in good agreement.*

cell has a diameter as large as 25 mm, but if we use a beam of 1 mm diameter we will be able to use the successor to the current PA cell (not published), which has a sensitivity of two orders of magnitude greater, increasing the detectivity to $\approx$ 10 ppb. Combining this with an optimized coating on both QCL facets, the detectivity will be approximately 1 ppb.

3.4.2 Quantitative Discussion of the role of the AR coating

In this section we will compare performance of the cases with and without an AR coating and and give an analysis of the lasing modes and linewidth on the example of the 11.1 μm laser.

Without an AR coating, the intra-cavity facet has an as-cleaved refectivity of 28%. If used in the EC setup, the useful power in the grating-selected modes is severely limited by the dynamic range in which the laser can be driven without oscillation on the chip's own Fabry-Perot modes. For a 5-mm QCL from the same wafer as the QCL used in the previous section, in the center of the gain region, the threshold current for lasing on the grating-selected mode was 1.757 A, while the threshold for the FP modes was 1.802 A. This resulted in a maximum power in the grating-selected modes of no more than 380 μW. Owing to the limitation on the drive current, the tuning range is also limited, in this case to about 36 cm^{-1}, which is 4% of the center wavelength.

The 5-mm QCL used in the previous section (which had a slightly higher waveguide loss than the 6-mm QCL), before coating had a threshold current for the FP modes of 1.96 A. When coated with an AR coating of 9.0% reflectivity, the threshold rose to 2.28 A, while the threshold for the EC-selected modes was lowered to 1.86 A. While the uncoated QCL could be driven 2.6% above its lasing threshold in the EC, the coated one can now be driven up to 17.2% above its threshold, which leads to a higher output power and a broader tuning range. In the center of the gain region with a pulse duration of 100 ns and a repetition rate of 200 kHz, 2 mW average power on the grating-selected modes was reached and the tuning

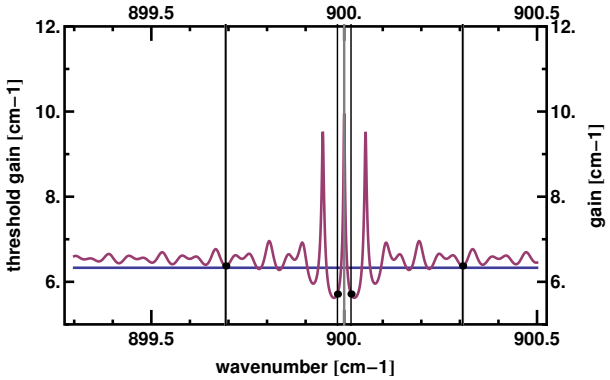
range almost doubled to 67 cm^{-1}, which equals 7.5% of the center wavelength.

As discussed in Section 1.5.6, another drawback of an imperfect AR coating is associated with coupled-cavity effects that lead to intensity fluctuations depending on whether or not the center of the grating-selected bandwidth (the *grating* or *Littrow wavenumber*) coincides with an FP mode of the QCL chip. For the extreme case of not using an AR coating at all, the intensity periodically vanishes with a period of $1/2nl \approx 0.308$ cm^{-1} (for the 5 mm QCL), i.e. the period of the chip's FP modes. Since the Littrow reflection of the grating has a bandwidth of $\delta\nu \approx \nu/N = 0.166$ cm^{-1}, where $\nu = 900$ cm^{-1} and $N \approx 6800$ is the number of illuminated grooves on the grating, it selects as many as 2-3 external cavity modes, spaced at ≈ 0.046 cm^{-1}. Closer analysis of the setup with the model given in [22] reveals that all of these modes get strongly suppressed when the grating bandwidth center is detuned from the FP modes. Figures 3.14 a) and b) show the threshold gain envelope function (wavy line) as a function of wavenumber for two different settings of grating angle. Figure 3.14 a) shows the threshold gain for the case where the center of the grating feedback bandwidth is tuned onto an FP mode (drawn as thick gray vertical lines) at 900 cm^{-1} and b) where it is centered between two modes at 900 cm^{-1} and 899.846 cm^{-1}. The black vertical lines indicate allowed modes of the external cavity, i.e. the phase condition is satisfied on a round trip including the free space path toward the grating. The horizontal line shows the gain corresponding to the pump current the laser is driven at. The intersecting points of the black vertical lines with the threshold gain envelope indicate the thresholds for those particular modes. All modes that have thresholds below the actual gain will lase (in pulsed mode, i.e. without full mode competition or gain saturation). The lower these points lie, i.e. the greater the difference between gain and threshold gain, the stronger the mode will lase. In Fig. 3.14 b) it is apparent that the modes (black lines) are sufficiently strongly suppressed for the laser to shut off. The greater the distance between FP modes, i.e. the shorter the QCL stripe, the wider the tuning gap, i.e. the range in which there is no intensity.

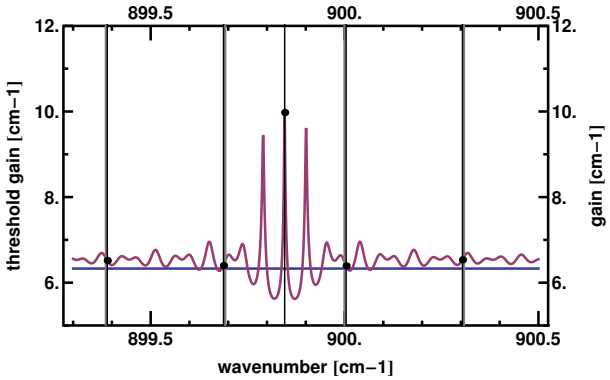
Preliminary results show, that on a 1-cm laser the power does not completely vanish between FP modes due to the high density of the FP modes whose spacing now is then exceeded by the wavelength chirp during the pulse.

For the AR coated stripe, Fig. 3.15 shows the output power as a function of the grating-selected wavenumber (center of bandwidth). Between FP modes, the laser does not shut off, but the power decreases to about 26% of its next maximum value. The oscillations have a period of 0.308 cm^{-1} which is equal to $1/2nl$ with n =3.25 and l = 0.5 cm. Also, the maximum power very slightly oscillates with a period that is equal to about 4 periods of the FP modes, as is to be seen in Fig. 3.16, which is a magnification of some of the oscillations in Fig. 3.15.

The first of these results is explained in Fig. 3.17 using the same model used for Fig. 3.14, the second is explained in more detail in the next section. Figures 3.17 a) and c) show cases where the grating is tuned to FP modes of the QCL at 900 cm^{-1} and at 899.692 cm^{-1} yielding power maxima, and b) where it is tuned halfway between these two modes, resulting in a power minimum. The EC modes are marked with black lines and the FP modes with



(a) *Grating is tuned to FP modes of the QCL at 900 cm^{-1}*



(b) *Grating is tuned midway between FP modes to 899.846 cm^{-1}.*

Figure 3.14 *Analysis of the lasing longitudinal modes of the uncoated laser in the external cavity. The wavy line indicates the threshold gain as a function of wavenumber when the grating is tuned to a) 900 cm^{-1} and b) 899.846 cm^{-1}. Fabry-Perot modes of the chip are drawn as gray lines at 899.692 cm^{-1}, 900.00 cm^{-1}, and 900.308 cm^{-1}. Black vertical lines indicate allowed longitudinal modes of the entire cavity, where the phase condition is satisfied not just within the chip. The horizontal line indicates the gain at which the laser is pumped, since it is driven with short pulses, gain clamping can be ignored to first approximation. The intersecting point of the wavy line with a black vertical line (marked with dots) is the threshold gain for that particular mode, if the crossing point is below the horizontal line, it will lase. The lower the threshold gain, the stronger this mode will lase. In a) there are two lasing modes, in b) the laser has shut off.The parameters for the QCL used in this model were: waveguide loss $\alpha_w = 4\ cm^{-1}$; gain in the limit of no amplifier saturation $\gamma_0(I) = 3.33 \times I[W]$; drive current, $I = 1.9\ A$, free space length $l = 9\ cm$, coupling constant between chip and free-space cavity $\eta = 0.9$, grating efficiency $R_G = 0.93$.*

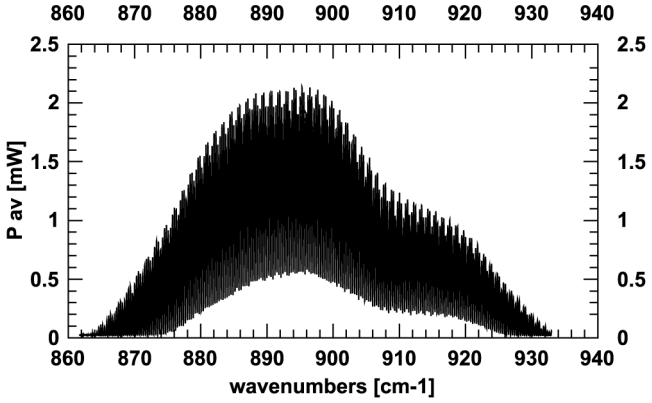


Figure 3.15 *Power as a function of grating-selected wavenumbers for the AR-coated QCL. Strong oscillations on the FP modes are visible, but the power does not shut off in between modes.*

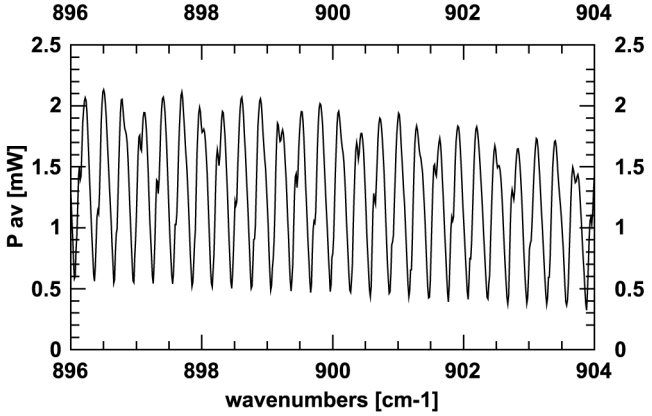
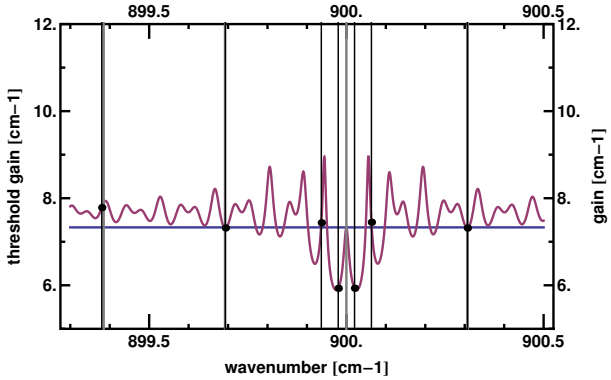
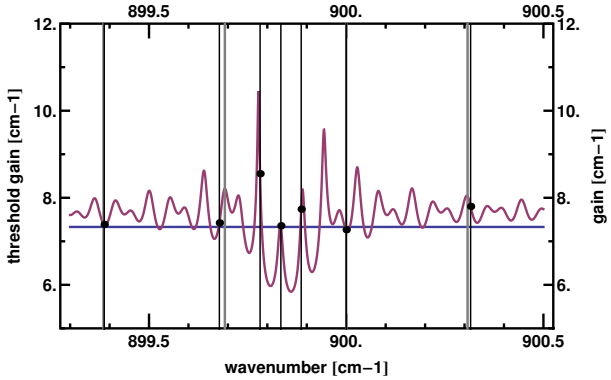


Figure 3.16 *Magnification of the curve plotted in Fig. 3.15 in an interval around 900 cm^{-1}. Clearly visible are the oscillation of the power maxima with a period of approximately 4 times the period of the basic oscillation as well as the sharpness of the power minima, both effects are explained in detail in Section 3.4.3*

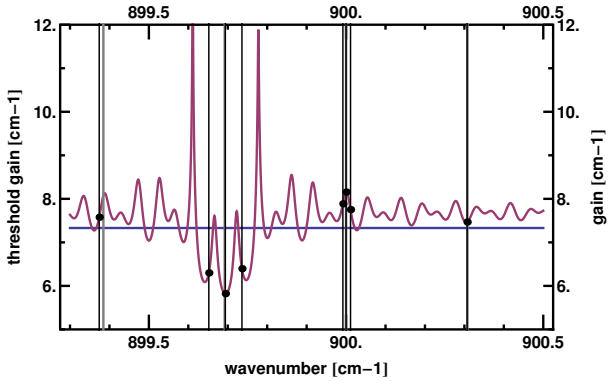
.



(a) *Grating is tuned to FP mode of the QCL at 900 cm^{-1}*



(b) *Grating is tuned midway between FP modes.*



(c) *Grating is tuned to FP mode of the QCL at 899.692 cm^{-1}*

Figure 3.17 *Threshold gain envelope (wavy line), Pump gain (horizontal line), FP-modes of chip (black vertical lines), EC modes (thick gray lines) of the coated QCL. a) and c) result in power maxima, b) results in a minimum. The parameters for the QCL used in this model were: waveguide loss $\alpha_w = 4\ cm^{-1}$; gain in the limit of no amplifier saturation $\gamma_0(I) = 3.33 \times I[W]$; drive current, $I = 2.2\ A$, free space length $l = 9\ cm$, coupling constant between chip and free-space cavity $\eta = 0.9$, grating efficiency $R_G = 0.93$.*

thicker grey lines. In these plots, the intersections of the allowed modes and the loss curve are also marked by dots for clarity. The higher the loss of a mode, the lower the power in this given mode for a certain Littrow wavenumber ν_g given by the grating angle.

The difference between cases a) and c) in this simulation is that in a), the effective free-space cavity length between the intra-cavity facet and the grating does not allow for a standing wave at exactly 900 cm^{-1} in a) (because of the phase jump of π upon reflection from the grating), but in c) at 899.692 cm^{-1} it does. Thus in a) the effective EC reflectivity decreases at exactly 900 cm^{-1}, decreasing laser action there. The next EC modes are very close though, and laser action is strong on both of these. The output power of each mode is proportional to the difference between gain and threshold gain (for short pulses, i.e. disregarding mode competition).

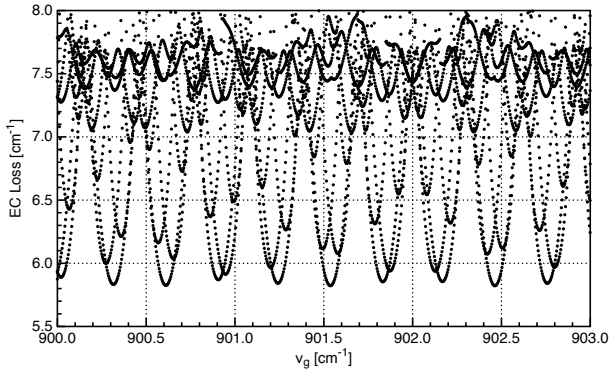


Figure 3.18 *Loss values of all the allowed EC modes as a function of grating wavenumber. Grating pivot point 2.5mm from optical axis.*

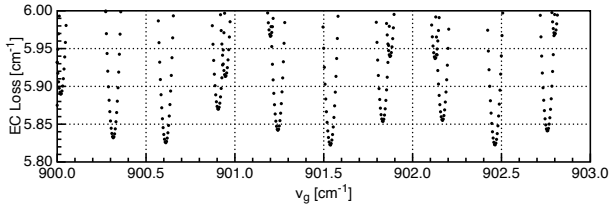


Figure 3.19 *Zoom-in on loss values of all the allowed EC modes as a function of grating wavenumber. Grating pivot point 2.5mm from optical axis.*

3.4.3 Quantitative Discussion of mode-hopping for the AR-coated laser

It is interesting to gather the loss values of the allowed modes (the ordinate value of the black dots in Figs. 3.17) as a function of grating wavenumber ν_g. This is plotted for the case of this particular EC in Fig. 3.18. As can be seen, the loss value of the EC overall (the minimum loss for a given ν_g), oscillates with the period of the FP modes, 0.308 cm^{-1}. The plot of dots can be seen as a collection of parabolas each of which corresponds to a particular EC mode that, as the grating is tuned, comes and goes into and out of resonance (at the loss minimum

of a parabola). Wherever two parabolas cross, the laser has two allowed modes with exactly the same loss. If the grating is tuned to a point where the minimum loss value is a crossing point (e.g. at 900.94 cm^{-1}), the laser has two strongest modes. This results in multimode oscillation even in the case of CW operation. In the case where one mode is much stronger then the two adjacent modes, e.g. at 901.54 cm^{-1}, the laser has the best possible side-mode suppression ratio. There are two more interesting aspects to point out about this graph. First is that the overall loss of the laser has parabolic minima and sharp maxima (where two parabolas meet in the middle between FP modes). This behavior can be found in the power envelope Fig. 3.16, since a loss minimum translates proportionally into a power maximum. The second is that the overall loss minimum slightly oscillates over several modes, which can be seen more clearly in Fig. 3.19, which is a zoom-in on the minima of Fig. 3.18. This behavior can also be directly verified in the slight mode-to-mode oscillation of the maximum power in Fig. 3.16. The period of this super-oscillation is a function of the distance of the grating's pivot point from the center of the optical axis. In order to arrive at a period of approximately 4 times the FP mode period (as is the case in Fig. 3.16), a pivot distance of approximately 2.5 mm is found and used for the simulation of Fig. 3.18, which is reasonable given the experimental setup.

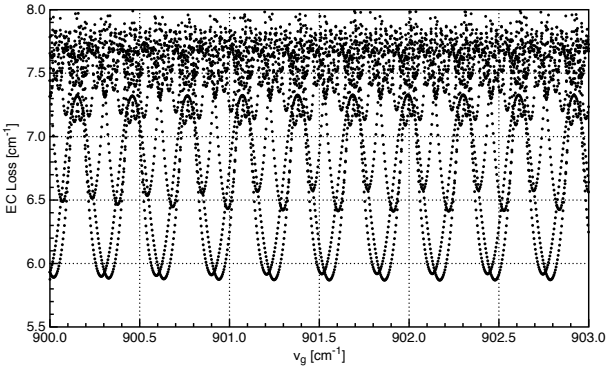


Figure 3.20 *Loss values of all the allowed EC modes as a function of grating wavenumber. Grating pivot point 60mm from optical axis.*

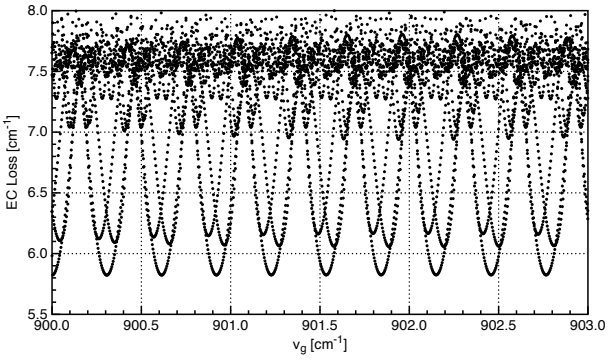


Figure 3.21 *Loss values of all the allowed EC modes as a function of grating wavenumber. Grating pivot point 60 mm from optical axis. Cavity length corrected to favor one mode at a time.*

Repeating the simulation with a pivot point exactly 60 mm away from the optical axis,

we arrive at the modal behavior given in Fig. 3.20, which is fully periodic, resulting in no super-oscillations. This is because now as the grating is tuned, the overall cavity length adjusts itself proportionally to the change in grating wavenumber (within a certain range, more correct pivot points can be found [106, 107]). Repeating the simulation again with a corrected starting cavity length to favor a single mode strongly with respect to the two adjacent modes, we arrive at the situation given in Fig. 3.21.

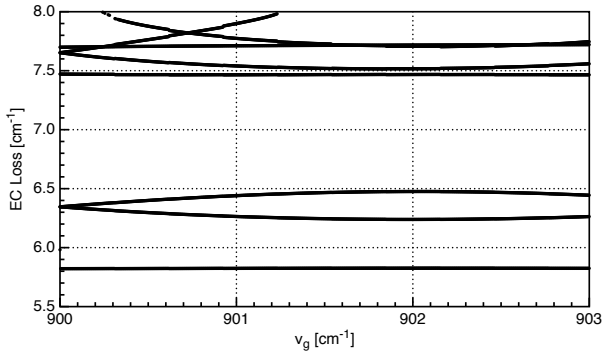


Figure 3.22 *Loss values of all the allowed EC modes as a function of grating wavenumber. Grating pivot point 60 mm from optical axis. Cavity length corrected to favor one mode at a time.*

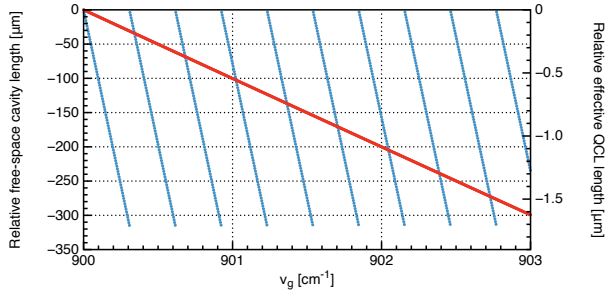


Figure 3.23 *Relative dynamic effective cavity lengths of the free-space section (red, left y-axis) and the QCL chip (blue, right y-axis) with respect to 90.00277 mm and 5 mm, respectively.*

Truly continuous tuning of a single mode can now be achieved by adjusting the temperature of the QCL dynamically in a way that the grating-selected mode always coincides with an FP mode. This could be done for instance by applying a saw tooth-like heating current to periodically vary the QCL's heatsink temperature and with it the refractive index and the effective optical length of the QCL waveguide as the grating is tuned. This of course is only of value with a CW QCL, which is why the experiment was not carried out. But the simulated behavior in analogy to Fig. 3.18 through Fig. 3.21 is plotted in Fig. 3.22. As is clear here, the mode-hops resulting from the FP oscillations have seized and the lowest-loss mode continuously moves along with the grating with a near-constant loss. The relative effective cavity lengths of the free-space section and the QCL chip are plotted in Fig. 3.23 with respect to 9.00277 mm and 5 mm, respectively.

The fact that the V-ECQCL setup resolves not only the FP mode oscillations, but also the super-oscillation in Fig. 3.16 that results from several different EC modes interacting, is a strong advantage of this setup, since only this allows the analysis made in this section be

experimentally verified even with a pulsed QCL. The enabling reason is that the collimating optic (the 1-inch parabolic mirror) is very large and with it the beam diameter so that the grating has a very high selectivity. Shrinking the beam diameter to 4-5 mm, as is done in the next section for reasons of mechanical stability, increases the grating bandwidth by a factor of 5 to 6. This consequently results in too many modes interfering with each other to make any reasonable interpretation of mode-hopping effects possible for a pulsed laser.

4

Stabilized Littrow External Cavity QCL with refractive optics and ultra-stiff structure (S-ECQCL)

4.1 Introduction

The V-ECQCL discussed in the last chapter has some great scientific advantages, namely it is a great testing facility for any semiconductor laser emitting anywhere between the terahertz and the ultraviolet regime, with an alignment that is independent of the actual power emitted by the laser. It has however one very severe drawback, which is mechanical sensitivity. This not only impedes its portability, but it even makes reproducibility of results between measurements volatile. This is mainly due to two reasons. One is the fact that very bulky optical components, i.e. the grating and the parabolic mirror, are mounted on standard kinematic mounts with many degrees of freedom and no way to lock them securely. The second stems from the fact that, due to their geometry, off-axis parabolic mirrors need comparatively large effective focal lengths. The mirror with a focal length of 1" in combination with QCL facet dimensions of approximately 10 μm limits the allowable angular misalignment of the mirror to approximately 1 arcmin as is illustrated in Fig. 4.1. Since the angle between incoming and outgoing beams is twice as large as the misalignment angle of a reflective element such as the grating, the tolerance is halved for the grating. And even with smaller misalignments, there will be significantly less coupling into the facet, which affects tuning range and output power.

The (time-averaged) linewidth of the ECQCL's emission is also directly correlated to the mechanical and thermal vibration of the setup because of the wavelength chirp induced by any angular motion of the grating. For instance, the absorption of the $C^{13}O_2$ isotopologue located at 2277 cm^{-1} has a linewidth of 0.2 cm^{-1} at 1 atmosphere and 300 K. But according

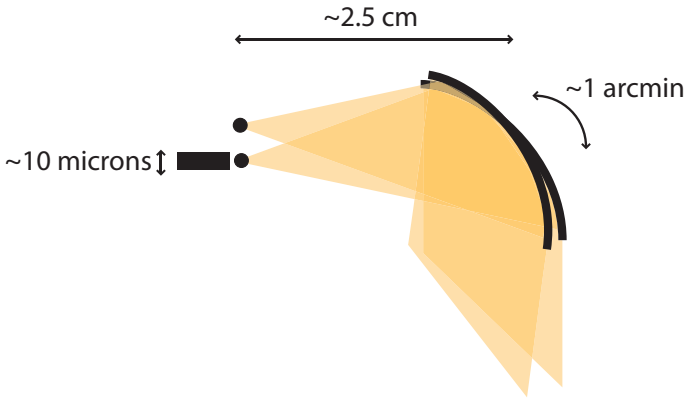


Figure 4.1 *The mirror with a focal length of 1" in combination with QCL facet dimensions of approximately 10 μm limits the allowable angular misalignment of the mirror to approximately 1 arcmin.*

to the Grating Equation, a 300 rules/mm Littrow grating is tilted by 16 arcsec to tune 0.2 cm^{-1} in this range. This means, that as soon as the angular vibration is of this order, the experimentally observable linewidth of this absorption line is limited by the vibration. For a grating mounted on a 3 cm actuator arm, this imposes a limit of approximately 2 μm of linear vibration. Although these values can be achieved on a well-damped optical table, this fact does limit the usefulness of the V-ECQCL to very carefully designed and well-protected environments. Also, these tolerances obviously scale down at lower temperature and pressure of the analyzed gas.

In this chapter, a Littrow ECQCL that was designed for extreme mechanical robustness and uses only custom-made mechanical parts is presented and its performance theoretically and experimentally investigated. This ECQCL has proven so reliable, that we have developed a fully operational, portable turn-key infrared spectrometer around it.

To understand the general design philosophy behind the structural engineering of the S-ECQCL, it is necessary to understand the way standard optical mounts as used in the V-ECQCL work. Figure 4.2 shows an illustration of the "Cone-V-Flat" principle used for kinematic optical mounts. The top part is mounted on the bottom part through its 3 hemispherical protrusions (blue) that sit in specially designed insets on the bottom part. The first (pink) is trihedral, the second is a V-groove (green), and the third is a flat (orange). Through this design, the top part touches the bottom part in exactly 6 well-defined points that constrain all the kinematic degrees of freedom (DOF) of one part relative to the other ("kinematic determinacy"). If the blue hemispheres are the tips of actuator screws (with the thread running through the top part), the relative positions of the top and bottom parts can be very accurately adjusted in three dimensions. In optical mounts, the top part is often the stationary one, and the bottom part holds the optical element. The two are pulled together by a spring.

The advantage is the ability to very accurately position one component relative to the other and to maintain this position over time. However, seen from a more practical standpoint, the configuration is extremely fragile. Consider the part holding the optical element. Instead of being fixed in space rigidly, i.e. bolted to the stationary part by the extreme pulling force

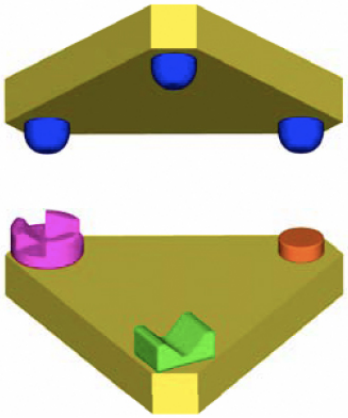


Figure 4.2 *Illustration of the "Cone-V-Flat" principle used for kinematic optical mounts. Image from [138].*

of massive steel bolts, it is held in 6 points with essentially no cross section and pulled into place by a relatively fine spring.

For the new design, the alignment and operation, as well as the tolerances of each DOF, have been carefully considered while developing each individual component. This resulted in a configuration where all DOFs of all parts are perfectly frozen after alignment except for the one degree necessary for operation, which is the Littrow angle.

Instead of adhering to complete kinematic determinacy, the strategy was to fully decouple alignment and fixation of each component. This was done as follows. As one component is actuated during alignment, it slides against a static component along a large polished area. After alignment, the parts are bolted together rigidly. The pulling force is now normal to the direction of actuation, thereby decoupling the two processes. This is illustrated for the example of the adjustment of the grating mount in Fig. 4.3. Acknowledging the fact that not all degrees of freedom require ultra-fine adjustment, but can be done by hand instead, the number of moving parts and the complexity of the design was further reduced by saving the actuation mechanism.

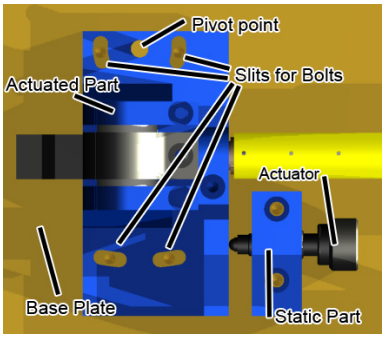


Figure 4.3 *Illustration of the decoupling of alignment and fixation for the example of the adjustment of the grating mount. The parts slide on polished surfaces and the fixation is done by using strong bolts that slide in curved slits before tightening.*

Also, all components were scaled down and designed in as compact a fashion as possible for maximum rigidity and stiffness by reducing internal leverage. This strategy raises the

mechanical resonance frequency and with it, it lowers the amplitude of any internal vibration.

Along with scaling down the components, the parabolic mirror with a focal distance of 1" was replaced by a miniature aspheric lens with a focal distance of 0.72 mm, reducing the sensitivity to angular vibration by a factor of 35. For a size comparison of the two, refer to Fig. 4.4.

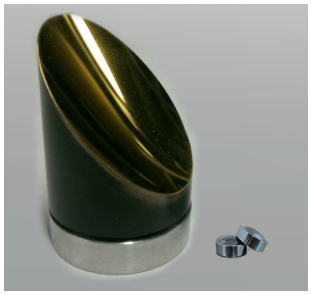


Figure 4.4 *For a size comparison of the parabolic mirror with a focal distance of 1" and the miniature aspheric lens with a focal distance of 0.72 mm.*

The only degree of freedom required for operation is the tuning angle. To eliminate wobble and maximize repeatability, no ball or plain bearings were used, but a pre-stressed, stiff, crossed double-flexure bearing as pictured in Fig. 4.5 was used instead. This reduces the repeatability error and wobble to a scale of nanometers in spite of the still relatively large actuated mass of the grating and grating arm. To further increase repeatability, the actuator head of the dc-servo motor touches the grating arm on a polished sapphire plate.

Figure 4.5 *Left: Pre-stressed, stiff, crossed double-flexure bearing. right: Cut-out for illustration of the working principle. Source: [139]*

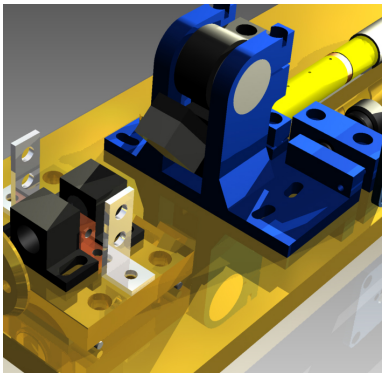


Figure 4.6 *CAD rendering of the S-ECQCL with the mounted QCL.*

Fig. 4.6 shows a CAD rendering of the S-ECQCL with the QCL mounted on the red sub-

mount surrounded by the black lens mounts all bolted onto a TEC-cooled sub-baseplate. The grating mount is shown in blue and the grating arm is shown in black. The grating angle that is normal to the tuning angle is adjusted by adjusting the entire grating mount.

4.2 Hardware

4.2.1 Modules

The full spectrometer (without the sample to be investigated) as photographed in Fig. 4.7, consists of five encapsulated modules. These are the sealed and cooled laser head, the controller housing, a hand-held controller for the QCL and TEC, the external intensity detector, and the computer that acquires and analyzes the spectra.

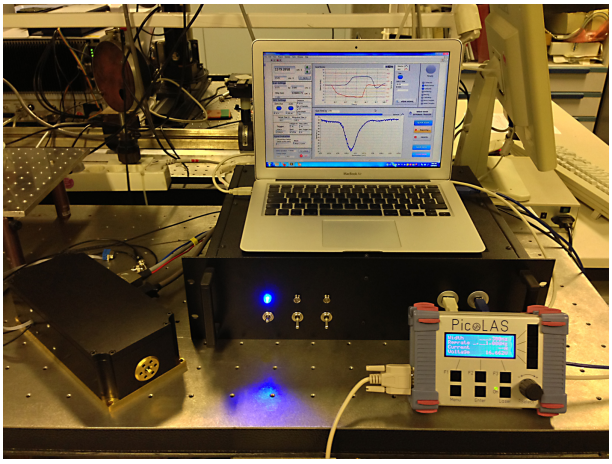


Figure 4.7 *Photograph of the full S-ECQCL spectrometer (without the sample to be investigated).*

4.2.1.1 Laser Head

The laser head contains the entire S-ECQCL as well as the QCL driver and TEC driver circuit boards, along with a PT100 temperature sensor. It is hermetically sealed to maintain a vacuum or contain the purge gas and is water-cooled, which is described in more detail in Section 4.10. The collimated laser beam exits through a wedged and AR-coated germanium window. The laser head is electrically connected with one mixed D-Sub connector 17W2 to supply the QCL driver and TEC driver with power (48 V on high-current connectors) and control signals, as well as an HD15 D-Sub connector to supply the dc servo motor with signals from the motor controller. It has outside dimensions of $11 \times 26 \times 8\,\mathrm{cm}^3$. Photographs of the closed and opened laser head are provided in Fig. 4.8.

The QCL driver in use is a combination of the LDP-V 50-100 V3 pulse driver and PLCS-21 frequency generator by PicoLAS GmbH. It can generate pulses of 3-50 A with a maximum output voltage of 100 V with a repetition rate of 1 Hz-2.4 MHz and pulse durations of 10 ns-10 μs. It has a current monitor of 20 A/V into 50 Ω.

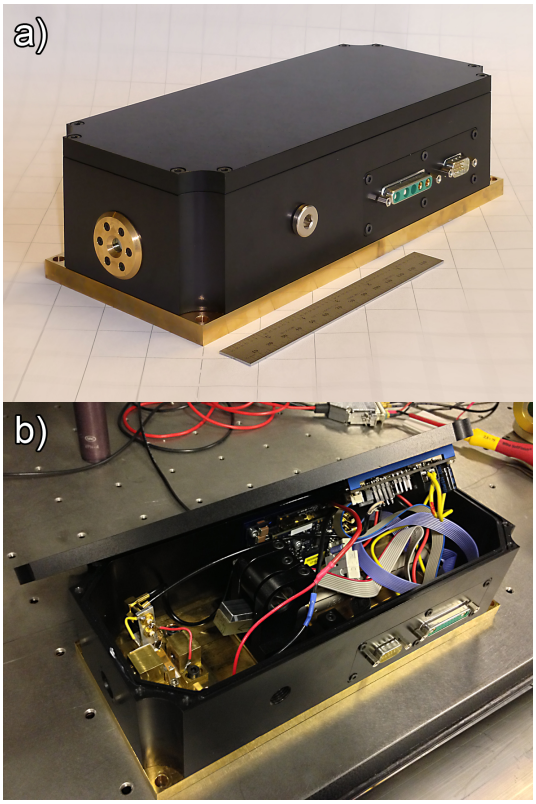


Figure 4.8 *Photographs of the closed (a) and opened (b) laser head without external cables.*

The TEC driver is a PL-TEC 2-1024 also by PicoLAS GmbH. It can heat or cool with a current of 10 A and a maximum power of 240 W. It provides closed-loop PID-controlled temperature stabilization with a PT100 temperature sensor to within 0.001 K.

Actuation is performed by the vacuum-rated dc servo motor Z812V by Thorlabs with 12 mm travel. This motor has a minimum repeatable incremental movement of 0.2 μm, which is approximately 0.8 arcsec on a 5-cm arm and a homing repeatability $\pm 1.0\,\mu$m which is approximately 4 arcsec on a 5-cm arm.

4.2.1.2 Controller Housing

The controller housing contains the larger electronics. These are the 48-V power source that supplies the QCL driver and TEC driver circuit boards inside the laser head, the T-Cube motor controller by Thorlabs, and the Picoscope 2204 digital oscillator that serves both as a fast analog-to-digital converter (ADC) for the detector, the pulse current monitor, or the external master trigger, as well as a generator for a current pulse that signals the finalization of a motor step to an external synchronizer.

It has three flip switches on the front with LED indicators to switch the power, cooling and laser emission. It has two USB connectors on the front to connect the motor controller and the digital oscilloscope to the computer, as well as two DE9 D-Sub connectors to connect to the hand-held device that controls the QCL and TEC drivers. On the back, it has one 230 V power connector, and three BNC connectors to connect the ADC to the detector and

the current monitor or external trigger, and the signal generator to other externally-triggered devices. The controller housing is designed to fit into a 19-inch rack. Photographs of the closed and opened controller housing are provided in Fig. 4.9. All electronics are grounded and fused.

Figure 4.9 *Photographs of the closed (a) and opened (b) controller housing without external cables.*

4.2.1.3 Hand-Held Controller

Through the hand-held controller, the parameters for the QCL driver and TEC are set. For the QCL driver, these are pulse current, pulse duration and repetition rate, along with several safety parameters such as shut-off current and voltage and over-heat protection. For the TEC driver, these are the set temperature and PID parameters for closed-loop temperature control. The hand-held controller does not necessarily need to be connected at all times, since once set, the parameter values are stored inside the drivers on non-volatile memory. Output from the drivers is switched on and off using the flip switches on the 19-inch controller housing.

4.2.1.4 Wiring and Connectors

The Laser head is connected to the Controller Housing with a D-Sub HD15, a mixed D-sub 17W2 (2x high current, 15x signal). The first is fed through to the motor controller, the second is split up inside the controller housing. The high current wires are connected to the 48 V power supply that is used to supply the TEC controller and Laser Driver within the Laser Head with power. The signal connections are internally wired to the two DE9 connectors on the front of the Controller Housing to connect to the Hand-Held Controller to control the operation parameters of the TEC controller and the Laser driver circuits.

The Laser head has two more BNC connectors, a laser pulse trigger input and a signal monitor output, the latter of which can be connected to the Controller Housing.

The Controller Housing has two BNC inputs and a BNC output, internally wired to the ADC/signal generator. The inputs are for the external detector, the pulse monitor, or alternatively the external motor step trigger. The output generates a pulse at the end of each motor step.

The ADC and the motor controller are internally wired to two USB connectors on the front of the Controller Housing to connect to the computer.

This highly versatile setup allows integration into very complex systems, e.g. pump-probe experiments, where an external clock fires a trigger to make two different lasers fire a pulse at a predefined time delay. This configuration can also be combined with wavelength tuning of the ECQCL through the external step trigger. The signal from the detector can then either be evaluated by the integrated computer or a different, external computer.

4.2.2 Cooling

4.2.2.1 Thermo-electric Cooler

To maximize emission power, the QCL is cooled using a 4 cm $\times$ 4 cm thermoelectric cooler (TEC) inside the Laser Head with 250 W maximum power and maximum $\Delta T = 68\,K$. The temperature is stabilized to 0.001 K using closed-loop PID control with a PT100 temperature sensor and the TEC controller.

To avoid condensation, the Laser Head is hermetically sealed and has a inlet and outlet gas valves. These can be used either to reduce the pressure on the inside, to fill it with a dry inert gas, or to purge the head continuously.

4.2.2.2 Heat Exchanger

To extract the large quantities of heat produced by the TEC under high load efficiently enough not to cause performance degradation of the TEC, heat sinking needs considerable attention. A water-cooled heat sink with 400 W/K on an area of 5 cm × 5 cm was consequently designed with a model developed for this purpose, which is discussed in this section.

A water-cooled heat sink consists of a block of metal in direct contact with the QCL's copper submount, with a hollow internal structure through which a water flow is forced. The choice of metal is copper for its large thermal conductivity of 400 W/(K m), its inherent resistance to corrosion, and relatively low cost.

When designing a heat sink, certain aspects of fluid dynamics have to be considered. When water flows inside a tube, the *no slip* boundary condition at the wall dictates that the outermost "layer" of a water flow is completely static due to wall friction. In *laminar flow*, the innermost layer flows fastest, and all concentric layers – due to friction from the neighboring layers – gradually move slower the closer they are to the wall. These layers do not mix and the increase in flow speed towards the center is relatively slow. For more efficient convective heat transport to take place, the heat needs to reach the faster-moving water layers by traversing the slower-moving layers first. This first heat flow, however, happens normal to the direction of water flow, therefore the only available mechanism for heat extraction through these first laminar layers is inefficient *heat conduction*. The slow laminar layers consequently act as a heat barrier.

The way to reduce the thickness of this barrier for a given tube geometry is to strongly increase the flow speed. This is to increase the shear stresses between the laminar layers to the point of breaking, thus creating a very thin residual laminar sublayer abruptly followed by much faster water flow with a chaotic flow pattern. This is the *turbulent* regime, in which the heat resistance of the structure is rapidly decreased.

The transition between laminar and turbulent flow regimes inside a pipe can be characterized with the use of the well-known dimensionless *Reynolds number* [140]

$$R_e = \frac{QD_H}{\nu A}, \tag{4.1}$$

where Q is the volumetric flow rate, D_H the hydraulic diameter of the pipe, ν is the kinematic viscosity (of water in this case), and A is the cross sectional area of the tube. D_H is equal to the diameter for ducts with circular cross section. For ducts with rectangular cross section, where a and b are the height and width, $D_H = \frac{2ab}{a+b}$. Thus at constant flow rate (not constant pressure) thinner tubes lead to larger Reynolds numbers because of increased particle velocity. Experience shows, that if the Reynolds number is greater than approximately 4000, the flow is turbulent.

Thinner tubes, however, also have a larger pressure drop across them. When designing ducts across the heat sink, the pressure drop needs to be accounted for, since the flow rate drops to zero if the pressure drop across the structure is equal to the pressure supplied by the cooling water recirculator.

The pressure drop ΔP_1 is given by the *Darcy-Weisbach equation* [140]

$$\Delta P_1 = f \frac{\rho v^2}{2} \frac{l}{D_H} \tag{4.2}$$

where l the length of the duct and v the speed of the coolant. f is the *Darcy-Weisbach* or *Moody friction factor* that depends on the Reynolds number and the relative pipe wall roughness, according to the implicit *Colebrook equation* [140] (valid for turbulent flow)

$$\frac{1}{\sqrt{f}} = -2.0 \log_{10} \left(\frac{\epsilon/D_H}{3.7} + \frac{2.51}{R_e \sqrt{f}} \right), \tag{4.3}$$

where ϵ/D_H is the relative pipe roughness, where ϵ is the length scale of the roughness. For laminar flow $f = 64/R_e$, which is an analytic results from the *Hagen-Poiseuille Equation*. Both expressions are tabulated in the so-called *Moody chart* [141].

Normally, heat transfer even in turbulent regimes is limited by the thickness of the residual laminar sublayer of the flowing coolant that forms on the inner wall of the duct due to the "no-slip" boundary condition. This layer is not involved in the turbulent convection and heat transfer through it is still governed by conduction. The laminar sublayer thickness is

$$\delta_\nu \approx 11.6 \frac{\nu}{u_*}, \tag{4.4}$$

where $u_* = \sqrt{\frac{\tau}{\rho}}$ is the shear velocity and τ is the shear stress within the water flow and ρ is the density of the coolant. $\tau = \mu \frac{\partial u}{\partial y}\Big|_{\text{wall}}$, where u is the flow velocity and y the coordinate normal to the wall. The shear velocity can well be approximated by $u_* \approx \frac{v}{10}$, which indicates that greater overall flow velocities v also reduce the thickness of this detrimental boundary layer. It is also beneficial to make the walls rough in such a way that the length scale of the roughness is greater than the thickness of the laminar sublayer. This helps perturb the boundary layer and cause rough turbulent flow at the wall. However, a few things must be kept in mind when designing this roughness. As the roughness increases, it increases the friction of the water at the wall, thereby essentially laminarizing the overall flow. Further, if the roughness is too large, the exact geometry needs to be considered, but in most cases it is detrimental to heat transfer. In the case the roughness is made up of grooves that run parallel to the water flow, the water inside the grooves flows slower due to the great wetted perimeter inside these grooves causing great friction. If the roughness is made of grooves that run along the circumference of the channel, it leads to the formation of water pockets inside the grooves with nearly static circular motion, which thereby essentially do not contribute to convection. In both cases the water stream splits up into a fast-moving central channel and slow-moving layers on the order of the roughness scale by the walls.

The heat transfer coefficient through the pipe wall can be estimated for forced turbulent convection ($R_e \gtrsim 10000$), smooth walls, and *Prandtl numbers* between 0.7 and 120 (water: 9.5) from the *Dittus-Boelter equation* [140] to be

$$h = 0.023 \frac{k^{1-\tilde{n}} j^{0.8} c_p^{\tilde{n}}}{\mu^{0.8-\tilde{n}} d^{0.2}} \tag{4.5}$$

where k is the thermal conductivity of the coolant $\tilde{n} = 0.4$ for heating, $j = Q\rho/A$ is the mass flux, and c_p is the isobaric heat capacity. The overall heat exchange thus is

$$W = hA_c \tag{4.6}$$

where A_c is the cooling area, for a tube with rectangular cross section this is $A_c = 2(a + b)l$.

Thus the cooling power is directly proportional to the inner surface area of the tube, approximately proportional to the inverse of the cross-sectional area, and approximately proportional to the flow rate of the coolant. Obviously, if there are n parallel ducts, the flow rate through each duct scales as $Q \mapsto Q/n$ and the overall cooling area $A_c \mapsto nA_c$. The effect of multiple ducts approximately cancels out in the expression for the heat exchange, but it is important when balancing tight ducts for large flow velocities to reach high enough Reynolds numbers for turbulent flow against the overall pressure drop, so the flow rates can be sustained with a given coolant pump pressure. Besides improved heat transfer, turbulent flow is also a necessary condition for the above approximations to be valid.

If the length of the cooling ducts, as a parameter in the optimization process, is not to be bounded by the dimensions of the area to be cooled, it has to be folded or meandered within this area. However, each bend introduces further turbulences in the coolant and thus adds to the overall pressure drop. This can be calculated from [142, 143]

$$\Delta P_2 = f \frac{1}{2} \rho v^2 \frac{\pi R_b}{d} \frac{\theta}{180^\circ} + \frac{1}{2} k_b \rho v^2 \tag{4.7}$$

where f is the Moody friction factor, R_b is the bend radius, θ is the bend angle and k_b is the bend coefficient [144]. Thus if the area to be cooled has the dimensions $a_1 \times a_2$ if there are

$$m \approx l/a_1 \tag{4.8}$$

180°-bends and the overall pressure loss is

$$\Delta P = \Delta P_1 + m\Delta P_2 \tag{4.9}$$

Also, if n ducts (of the same length) have to fit within the cooled area and b is the horizontal dimension of a duct, it yields the constraint

$$nblx \leq a_1 a_2, \tag{4.10}$$

where x is the filling factor of the area, e.g. 1/2 or 2/3 to allow for walls (fins) between the ducts. The filling factor can best be adjusted iteratively, by acknowledging the fact that with a given filling factor the walls have a height/thickness ratio that can be chosen in such a way that the heat resistance of the copper fins approximately equals the heat resistance of the transfer through the fin-water boundary.

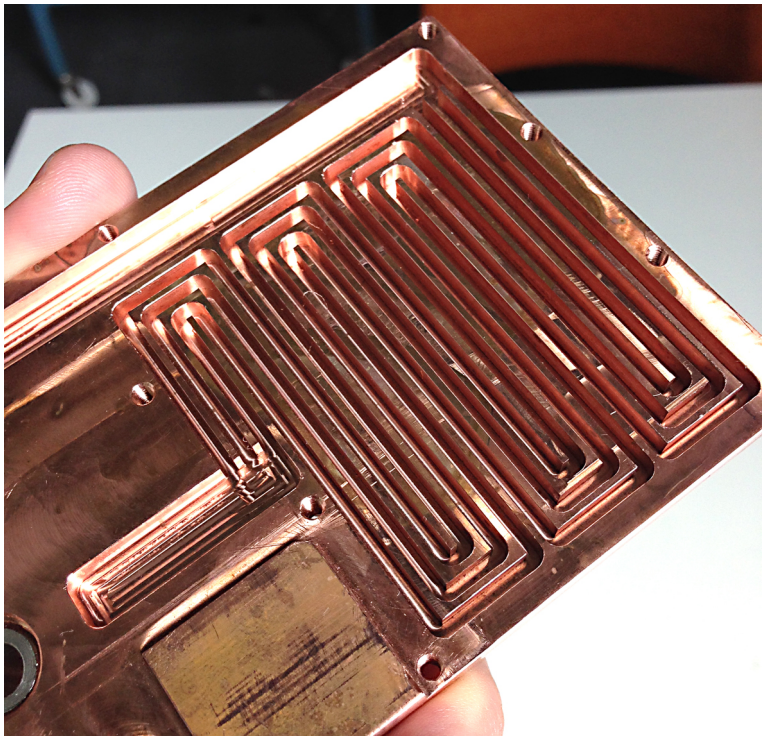


Figure 4.10 *Photograph of the heat exchanger as manufactured according to the design calculated in this section.*

For the given situation of designing the water-cooled heat sink for a 4 cm×4 cm TEC cooler the solution lies in maximizing Eq. (4.6) for a given volumetric flow $Q = 4$L/min and free parameters n, l, a, b, while obeying the conditions $R_e \geq 4000$, $\Delta P \leq 1$bar, Eq. (4.10) and $1mm \leq a \leq 4mm$ for the vertical duct dimension and $2mm \leq a \leq 10mm$ for the horizontal dimension and a filling factor of $x = 2/3$.

For obvious reasons the cooling area was extended to 5 cm×5 cm, and the result is $n = 3$, $a = 4\,mm$, $b = 2\,mm$, and $l = 28$ cm, which results in an effective cooling area of 83 cm^2 and a heat transfer of 104 W/K and a pressure drop of 0.25 bar.

A photograph of the heat exchanger as manufactured according to the design calculated in this section can be seen in Fig. 4.10. The overall thermal conductance measured is approximately 430 W/K, four times as high as calculated, but the pressure drop is also approximately 4 times as high, just under 1 bar, when the water flow rate is set experimentally to 4 L/min, the value underlying the calculation. The value was measured from a temperature drop of 0.45K at a heating power of 192 W using a PT100 temperature sensor. The pressure drop was measured using two manometers in the cooling water circuit, one before and one after the heat exchanger. Since the pressure drop and the thermal conductance increased by approximately the same value, some possible explanations are that the calculation is slightly wrong about the amount of turbulence the structure actually causes or that the mechanical parameters of wall roughness or corner sharpness are not as calculated. However, the heat exchanger is actually better than calculated, since it achieves the calculated conductance at much lower flow rates. It remains to be stated that we have not found a more efficient

conventional water-cooled heat exchanger in the literature or available for purchase.

4.3 Software

4.3.1 Introduction

This section provides an overview of the software written for the Spectrometer. It is not intended to be a complete programming documentation.

The software is a National Instruments (NI) LabView program written in NI's proprietary graphical programming language "G", in the form of an Asynchronous Queued State-Machine with Event-Driven Producer/Consumer Architecture. It handles the dc servo motor that actuates the Littrow grating and with it the ECQCL's emission wavelength, an analog-digital converter (ADC) that processes the information from the radiation detector, and a signal generator that is used to synchronize external devices and the stepper motor by sending clocking signals. It handles all the tasks of performing calibration, signal monitoring during the setup and run of experiments, to sampling scans, and exporting the measured data to files.

4.3.2 Queued State-Machine

A (finite) state machine (FSM) is a concept from automata theory, where an (abstract) machine can be in any one of a finite number of states, which contains all the information necessary for operation. The machine can transition to a different or the same state upon initiation of a triggering event or condition. The machine performs actions when entering or exiting a state, upon user input, or while transitioning between two states.

A queued state machine is an FSM that can store and retrieve data from an infinite-memory queue. It is a model of computation equivalent of a Turing machine, and thus lifts the fundamental computational limitations of a pure FSM, which stem from the finiteness of its memory. The queue is an ordered register, similar to a tape, that has a series of blocks printed on it. The blocks are packages (clusters) of information that generally contain the state to transition to and possibly a set of parameters or data to be processed in that state.

Event-driven Producer/Consumer architectures work by filling the queue from a so-called producer loop that is linked to the user interface (UI) of the software. Completely independent of this, a consumer loop that is the actual state machine, processes the commands in the queue in order. Typically, the last step during the execution of any state is to fill the next state to proceed to back into the queue, typically along with some processed information to be processed further. Thus the consumer loop has a producer-character as well.

State machines have great advantages over simple loop/sequential designs. The first one is stability. Per design, a state machine is always in a well-defined state. Thus a well-debugged state machine is never in an indeterminate state that might cause it to "freeze" due to entangled processing commands. This is particularly important when controlling physical

automation asynchronously, i.e. where timing can not be strictly dictated, but call-backs from the physical devices (motors, detectors, etc.) upon completion of tasks need to be waited for. The de-coupling of the different processes, such as user input acquisition, physical actuation, signal acquisition, and information processing, has a whole class of advantages of its own, i.e. highly responsive user interfaces, multi-threaded parallel processing that distributes computing resources better and saves time. Another advantage is that it facilitates overview during the development stage, leading by design to good programming practice. A state machine is generally much easier to debug than other designs, because of the strict modular nature of the states.

The queued state machine presented here does not remain in the same state for a longer period of time, but instead executes its operation in that state and immediately transitions to the next state. A distinguished state is the "waiting" state that performs no operation at all, except to fill the queue with another "waiting state command to proceed to next. Thus if no user entry is performed, the machine loops endlessly through the "waiting" state until the program is exited. Most other states finish by writing the "waiting" state back into the queue.

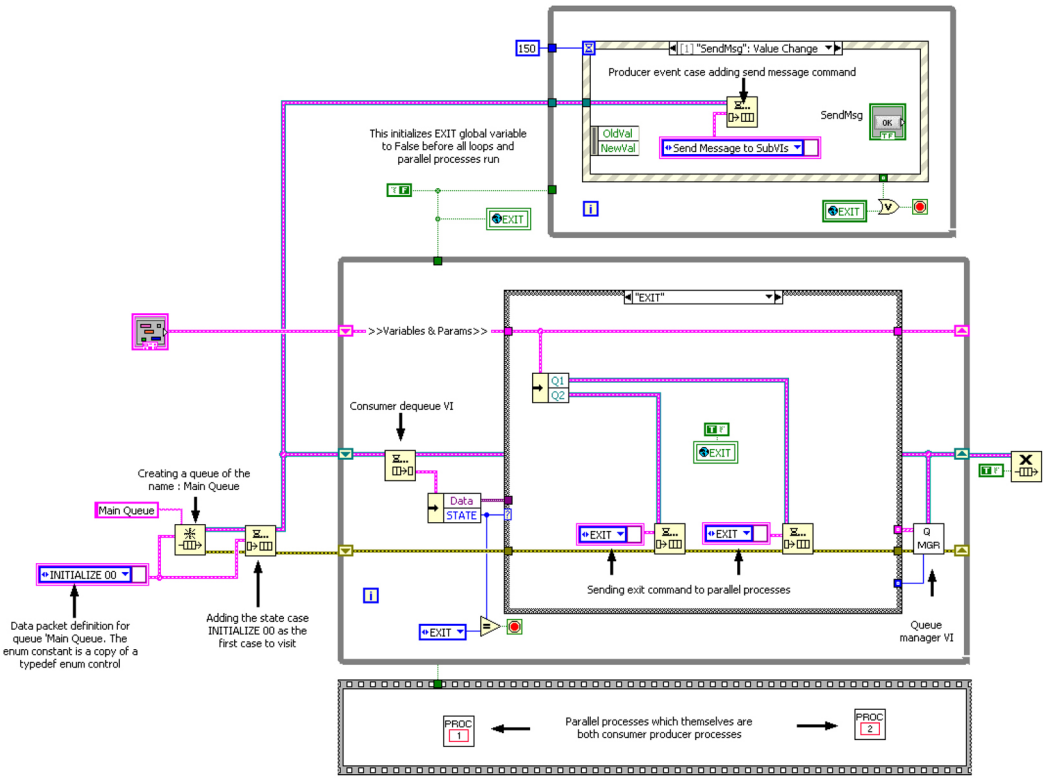


Figure 4.11 *Block diagram of a generic Event-Driven Queued State Machine with two external State Machines all of which used queued communication with each other.*

Figure 4.11 shows the block diagram of a generic state machine of this architecture. On the left side, the queue is defined and initialized with the first state "INITIALIZE 00". The state "INITIALIZE 00" performs initialization operations such as define variables, start devices, and so on. The queue is wired to two parallel loops, the top one is the producer loop and

the bottom one is the consumer loop. Inside the top loop is an event handler, that handles all the possible user inputs. This graphical representation is a stack showing only one case at a time. The case shown here is the “SendMsg:ValueChange” event, which is the event triggered when the user pushes the button "Send Message" on the front panel (not shown). In this case the “Send message to SubVis” state is entered into the queue. The bottom loop is the Consumer/State Machine loop. This dequeues the last element of the queue and fills it into a *case structure* for discrimination of the states. Here too, only a single state is shown at any given time, in the example here that is the “EXIT” state, which closes the program.

A typical information bundle inside the queue of the spectrometer is, e.g. [State: “go to position”; Data: “Position: 2300 cm^{-1}”]. The state machine then transfers to the state “go to position” in which the dc servo motor (that moves the Littrow grating’s angle) is called to go to the position specified. After the command is sent to the motor, the state machine reverts to the “waiting” state, which is initiated by filling the “waiting” state into the queue as the next state to proceed to. Writing the “waiting” waiting state back into the queue is the typical step performed upon completion of most states. Once the motor has reached its targeted position, it triggers an event “move complete”, which enters into the queue the command to update the UI to reflect the new current position of the motor and grating.

Different external processes such as signal acquisition can each have their own queues, making them queued state machines of their own. The different state machines then communicate with each other through adding information to the others’ queues. In fact, the shown “EXIT” state also closes the SubVIs by enqueuing the “EXIT” state into both of the SubVIs’ queues, Q1 and Q2, within the “Variables and Params” data cluster. This is an example of how the different state machines, the main one and the SubVIs, communicate through feeding information into one another’s queues.

4.3.3 The Front Panel (UI)

Figure 4.12 shows the UI or front panel of the software. It is divided into three main sections. On the left is the *Input Panel* for the experimental parameters, on the top right is the *Signal and Status Monitoring Panel*, on the bottom right is the *Spectrum Panel*.

4.3.3.1 Input Panel

The Input Panel is divided into six sections. The first section shows the current spectral position of the laser output in the large indicator, the bottom indicator can be filled with a new target wavenumber and confirmed with “Go To”. The up and down arrows are jog buttons to correct the output slightly upwards and downwards by pressing them repeatedly. The spectral information is converted into motor position and vice-versa using a calibration chart which is measured using an FTIR, stored in a file and read during software initialization.

The next section defines the scan parameters in terms of its upper and lower wavenumber limits and the resolution in terms of a step size between measurements.

The next section sets up the ADC for each measurement. There are two input channels,

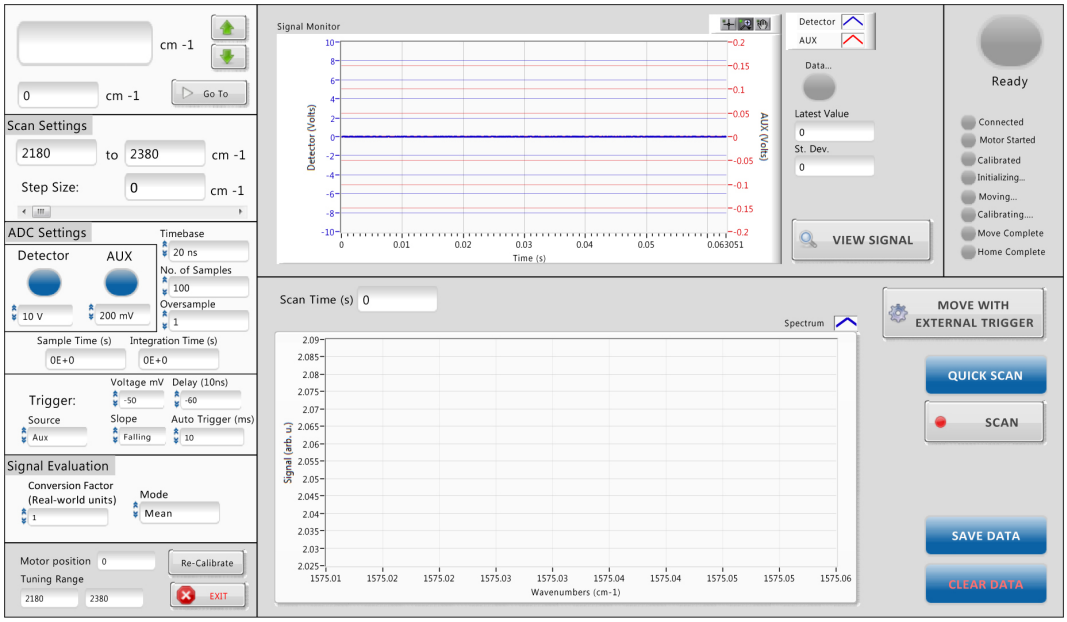


Figure 4.12 *The front panel of the Spectrometer*

one is for the signal detector ("Detector"), and the other is an auxiliary one ("AUX"), that can, for instance be used to monitor the pulse current driving the QCL as a sanity check for the experimental setup, or it can be used to react to an external trigger signal. This allows the spectrometer to be driven by a central clock of a larger experimental setup, for instance a pump-probe measurement where one clock is used to synchronize multiple lasers. For both, the resolution can be set. The "Timebase" and "No. of Samples" parameters determine the "Sample Time", and a number of pulses can be integrated over using the "Oversamle" input that then also determines the "Integration Time". The trigger of the ADC can be setup in the next panel.

The next section determines how a signal value is determined from an acquired pulse from the detector. First is the input of a calibration factor to compute units of electrical input to optical power, the second is the mode of how to collate the acquired signal to a single value. The options here are "mean", "peak height", and "area under pulse" (which integrates the pulse).

The last section contains the exit button of the program as well as some physical parameters of the motor as a sanity check since all the motor parameters discussed earlier are converted into spectral information.

4.3.3.2 Signal and Status Monitoring Panel

The Signal and Status Monitoring Panel is divided into two sections. The first shows a large signal monitor in the form of a two-input graph fed by the ADC and updated in real time. This greatly facilitates the setup and alignment of the expriment. This is because parameters to be set in the input section immediately reflect on the gathering of the signal on the next pulse, so timebase, resolution and trigger can be set like on a standard oscilloscope. The effect of varying the oversample parameter on the overall scan time can also be assessed immediately. Also, by maximizing or minimizing the detector signal, the detector and physical sample can be optimally aligned with respect to the ECQCL's probe beam or absorption lines of the sample can be located.

On the right of the graph, there is a status indicator that is on when data is currently gathered, and two numerical outputs showing the latest values of the collated measurement to be varied by varying the parameters in the "Signal Evaluation" panel.

Further to the right are a number of other status indicators showing, whether the motor is connected, started, calibrated, moving, whether the software is initializing, and also give off flashes when a movement is completed.

4.3.3.3 Spectrum Panel

The Spectrum Panel shows a large graph that updates in real time and contains the spectrum currently measured and keeps the spectrum visible until it is cleared using the button "Clear Data". The spectrum can be saved to file using the button "Save Data". The panel also contains buttons to initiate the scan ("Scan") and to perform an overview scan ("Quick Scan") with a resolution ("Step Size") automatically adjusted in the way that the quick scan always takes two minutes.

During the scanning process, a progress bar becomes visible and a button to cancel the scan.

There is also a button ("Move with external trigger") that sets the spectrometer into a waiting mode to perform a single step and a single measurement upon receiving a pulse from an external trigger on the "AUX" channel of the ADC. With a series of trigger pulses the spectrum is thus scanned and automatically finalized when it has reached the upper scan limit.

4.3.4 The Back End (Block Diagram)

Figure 4.13 shows the block diagram of the implementation of the generic state machine of Fig. 4.11 for this spectrometer. The producer loop only shows the "Cancel Scan" event in this figure. Altogether, there are there are 27 distinct events, whose event identifiers are pictured for overview in Fig. 4.14.

The consumer loop shows the "INITIALIZE 00" state. There are 36 distinct cases in the state machine (and a multitude of sub-cases), whose identifiers are once again shown for a rough overview in Fig. 4.15. Further, the program contains two queues, and one sub-VI for data acquisition. The motor is controlled using the ActiveX framework, and the DAC is controlled using imported C-libraries (.dll).

The choice has been made to neither show nor discuss cases and events in full here, since only printing the diagrams without comment would take up a lengthy 33 pages, without any interesting physics involved. Instead, one task – performing a Quick Scan – is picked out at random and operation of the software is explained for this example in Appendix E. This is relatively lengthy and serves to demonstrate the complexity involved with such a seemingly simple task of driving an ECQCL in a useful way.

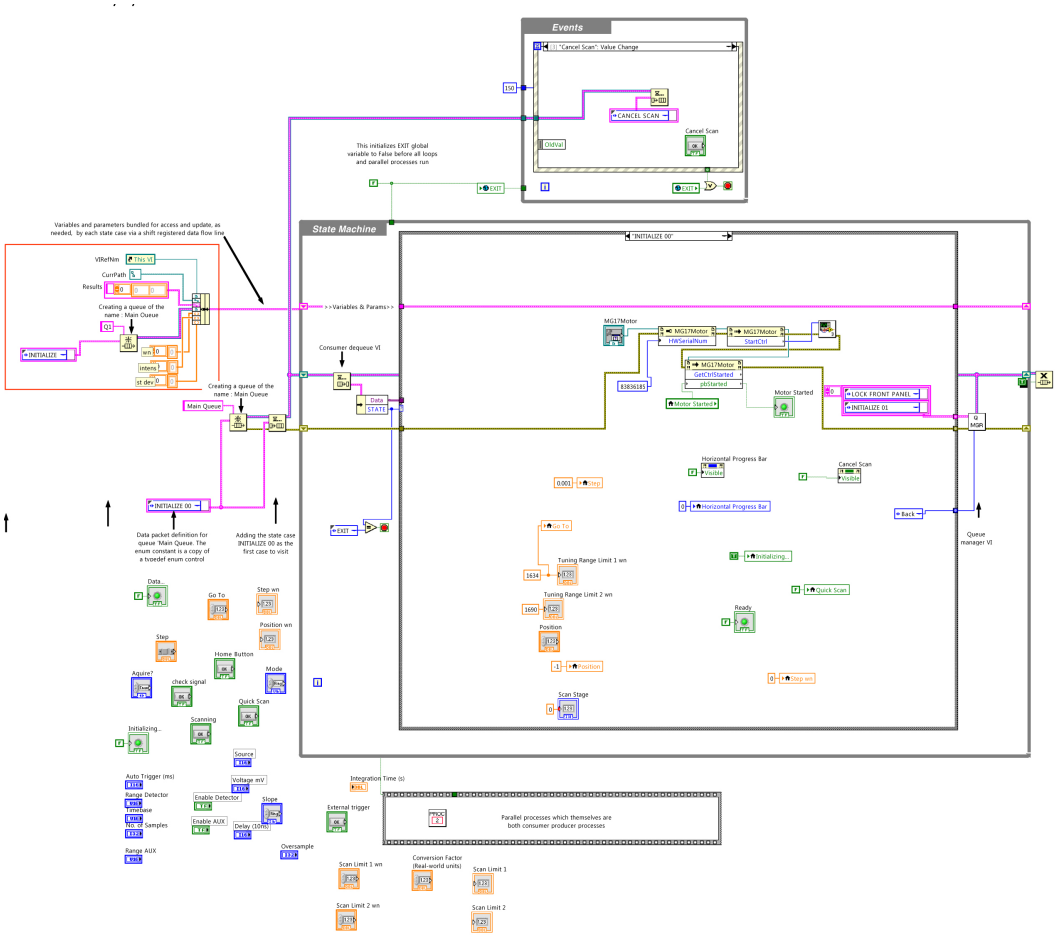


Figure 4.13 *Block diagram of the implementation of the generic state machine of Fig. 4.11 for this spectrometer.*

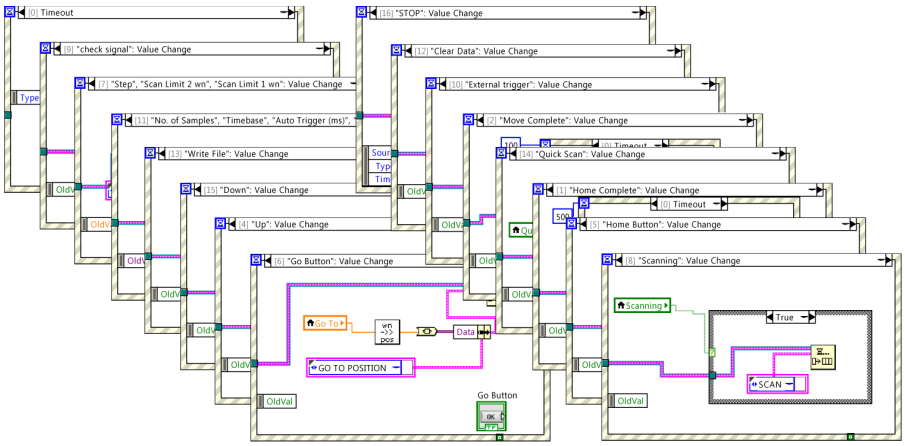


Figure 4.14 *Event identifiers of the state machine.*

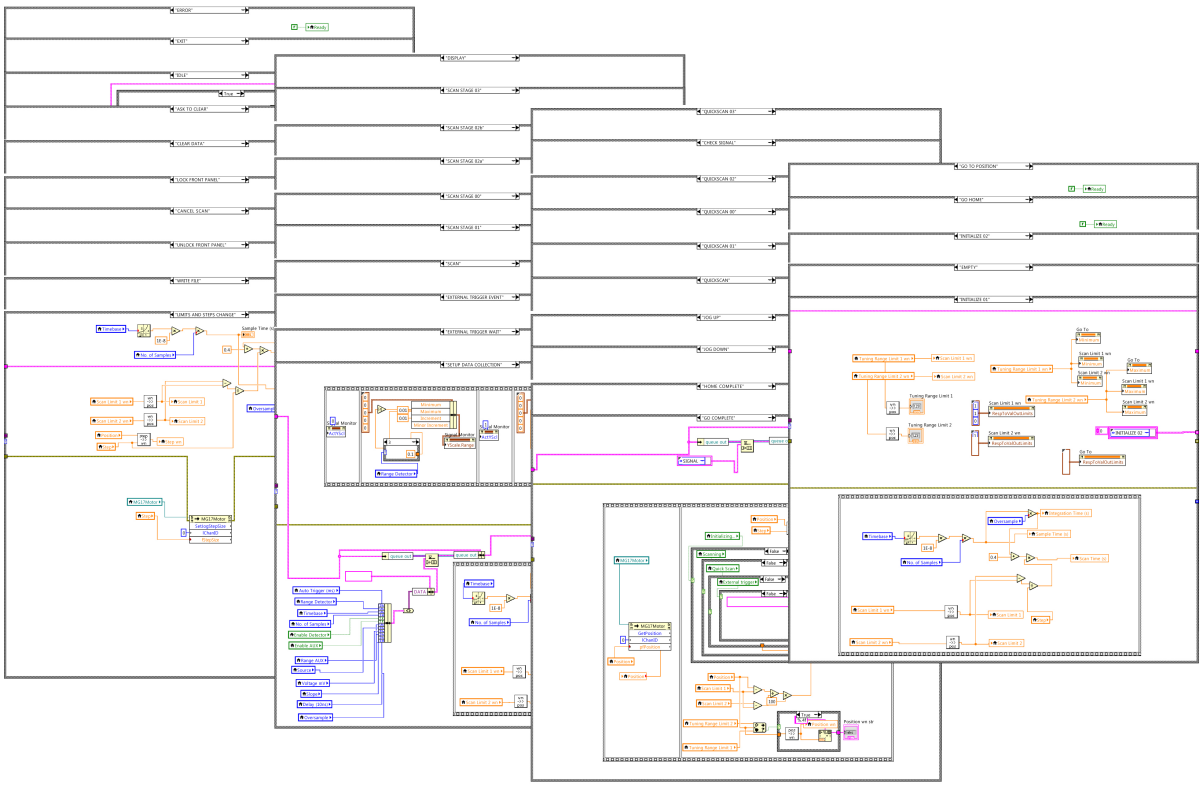


Figure 4.15 *Overview of the cases of the state machine.*

4.4 Performance

4.4.1 P-I curves and maximum power of purely grating-selected oscillation

Figure 4.16 shows the P-I curve (pulse power) of a QCL with the sample number HU2-0324 before application of an AR coating for two different temperatures and three different duty cycles. For the measurements of this section, the two temperatures are 21 °C, denoted as room temperature (RT) and -16 °C, denoted as TEC temperature or simply as "cooled". The QCL was driven with a pulse duration of 100 ns. The three different duty cycles are a) 0.1% (10 kHz repetition rate), which does not exhibit any effects of active region heating, b) 0.87% (87 kHz), up to which the average power of the uncooled QCL grows quite linearly with duty cycle, and c) 2% (200 kHz) as an extreme duty cycle where the temperature effect forces the emission into a rollover.

In Fig. 4.16 the thick lines represent the cooled QCL and the thin lines represent the uncooled QCL. The black lines are for 10 kHz, the red for 87 kHz, and the blue for 200 kHz. The threshold currents are noted in Table 4.1 along with the characteristic temperature T_0 for each duty cycle calculated from the change of threshold temperatures upon cooling. This is done by rearranging Eq. (1.38) to give

$$T_0 = \frac{T_1 - T_2}{\ln \frac{I_{th}(T_1)}{I_{th}(T_2)}}. \tag{4.11}$$

The characteristic temperatures range from 99 K to 106 K with no particular trend. It is safe to say that the spread comes from the experimental uncertainty, and is not a real effect.

As can be seen, the temperature effect of the duty cycle is more pronounced for the uncooled QCL than for the cooled QCL. For the cooled QCL, the decrease from 10-200 kHz is approximately -33% and for the uncooled it is approximately -55%, each measured at 7.5 A. This is because of the increased efficiency of the QCL when driven at lower temperatures.

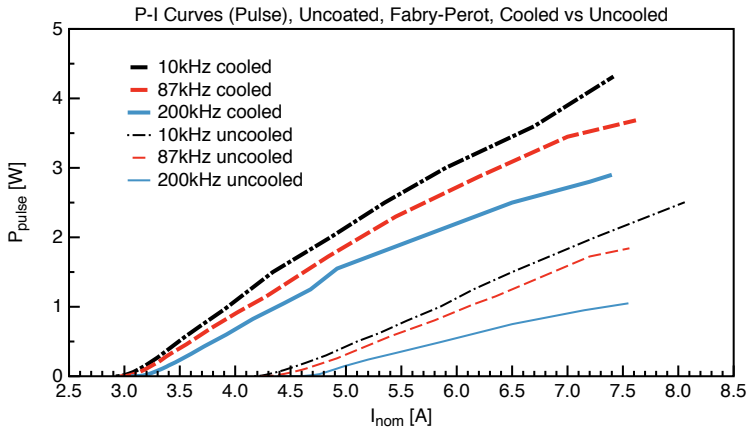


Figure 4.16 *P-I curves of the uncoated QCL at different temperatures and duty cycles.*

Table 4.1 *Threshold Currents of the uncoated QCL at different temperatures and duty cycles.*

Rep. Rate [kHz]	$I_{th}(\mathrm{FP}, \mathrm{Uncoated}, \mathrm{TEC})[A]$	$I_{th}(\mathrm{FP}, \mathrm{Uncoated}, \mathrm{RT})[A]$	$T_0[K]$
10	3.1	4.4	106
87	3.1	4.5	99
200	3.2	4.6	102

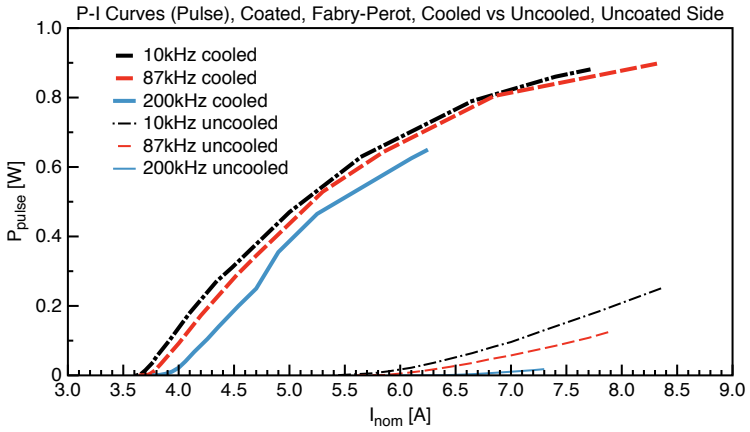


Figure 4.17 *P-I curves for FP oscillation of the coated QCL at different temperatures and duty cycles.*

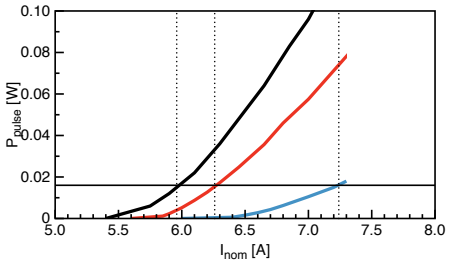


Figure 4.18 *The thresholds for the curves of the uncooled QCL in Fig. 4.17 and in Table 4.2 are tentatively assumed to be where the absolute pulse powers reach a value of 0.016 A.*

Table 4.2 *Threshold Currents for FP oscillation of the coated QCL at different temperatures and duty cycles (uncoated side).*

Rep. Rate [kHz]	$I_{th}(\mathrm{FP}, \mathrm{AR}, \mathrm{TEC})[A]$	$I_{th}(\mathrm{FP}, \mathrm{AR}, \mathrm{RT})[A]$	$T_0[K]$
10	3.65	5.69	75
87	3.75	6.26	72
200	3.95	7.24	61

Fig. 4.17 shows the P-I curves of the QCL running in FP mode after deposition of the AR coating onto one of its facets with the same color key as in Fig. 4.16. The AR coating has a residual reflectivity of approximately 1.5% as discussed in Section 2.5.6. An interesting point about this plot is that at the two different temperatures, the QCL shows very different character. While at TEC temperature, it shows behavior of a laser with a clearly defined threshold and rollover, at RT, it behaves like a super-luminescent diode with a P-I characteristic curved upwards with no clear threshold. This behavior has been discussed for this laser in detail in Section 2.5.3.

The threshold currents for the curves in Fig. 4.17 are given in Table 4.2. The thresholds for the uncooled curves are tentatively assumed to be where the absolute pulse powers reach a value of 0.016 A as can be seen in Fig. 4.18. This value is a reasonable guess estimate, as it is equal to the value of the output of the cooled coated and the uncoated laser at the threshold (which is determined as the intersection of the extrapolated slope with the current axis). If a threshold is defined this way, comparison with the thresholds of the cooled curves would result in an average $T_0 = 69$ K. But the exact value is of course not clearly defined. Though it is to be noted, that there seems to be an overall loss-dependence of the characteristic temperature, it seems to decrease with more lossy laser resonators. This needs further investigation.

Another very interesting point is the prominence of the rollover of the cooled QCL in Fig. 4.17, when comparing it to the cooled uncoated QCL in 4.16, which (at repetition rates up to 87 kHz) does not deviate from the straight slope up to 7.5 A. It is unlikely that the rollover is caused by overheating, since there are virtually no temperature effects from raising the repetition rate from 10-200 kHz in the cooled curves of Fig. 4.17. One possible explanation is the saturation of the upper laser level. At high pump rates the level gets filled strongly, while at lowered temperatures the non-radiative relaxation rate is decreased. But as the laser is AR coated, thus essentially an open resonator, with very low internal light levels, the stimulated emission rate is also suppressed. Therefore the carriers remain "stuck" in the upper level causing the emission to saturate at higher pump rates.

Fig. 4.19 shows the P-I curves of the coated QCL driven inside the EC, with the Littrow grating tuned to the position of the gain maximum. This plot also has the same color key as Fig. 4.16. The corresponding threshold currents and characteristic temperatures are given in Table 4.3.

Table 4.3 *Threshold Currents for EC oscillation (grating tuned to gain maximum) of the coated QCL at different temperatures and duty cycles.*

Rep. Rate [kHz]	$I_{th}(\mathrm{EC, AR, TEC})[A]$	$I_{th}(\mathrm{EC, AR, RT})[A]$	$T_0[K]$
10	2.9	4.4	89
87	2.95	4.55	85
200	3.1	4.6	94

When comparing the uncooled curves in Figs. 4.16 and 4.19, it is clear that the threshold current does not change at all (within the experimental uncertainty). The values for the

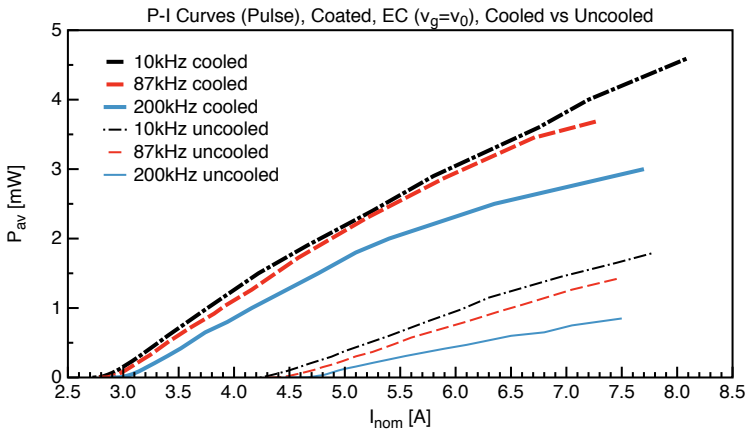


Figure 4.19 *P-I curves for EC oscillation in the gain max (coated QCL at different temperatures and duty cycles).*

different repetition rates remain in the range 4.4-4.6 A. From this it is concluded that the overall grating feedback is approximately equal to the feedback from the uncoated facet, which is 28%.

However, for the cooled QCL, there is a slight change from the range 2.9-3.1 A to 3.1-3.2 A which seems larger than the experimental uncertainty. This translates to a change of the average characteristic temperature from $T_0 = 102$ K for the uncoated QCL to $T_0 = 89$ K for the EC QCL. But T_0 is extremely sensitive to slight variations of the threshold current, so a small uncertainty in the threshold current translates to a very large uncertainty in T_0. It is therefore not quite clear if this is a real effect and needs further investigation.

Due to its high stability, the S-ECQCL is particularly useful for experimentally verifying the statements made in Sections 1.5.4 and 1.5.5 regarding the maximum tuning range and maximum output power of an EC driven in pure EC mode (i.e. without mixing FP oscillation into the emission). Since it has a very high level of reproducibility, it allows for the selective variation of individual experimental parameters without much fluctuation of the others.

As stated before, an interesting consequence of Eq. (1.46) is that, since the ratio $\frac{\alpha^{FP}}{\alpha^{EC}}$ is temperature-independent, the maximum achievable power with an ECQCL in pure EC operation is independent of temperature. This is remarkable, since the required pump rates change drastically with temperature. And since the duty cycle is also a temperature effect (because the current that heats the active region is proportional to the duty cycle), the maximum achievable power with an ECQCL in pure EC operation is also independent of duty cycle.

Figure 4.20 shows the P-I curves from Fig. 4.19 for the grating tuned to the gain maximum along with a section of the FP curves from Fig. 4.17 around the thresholds. The threshold currents of the FP curves from Table 4.2 are marked with vertical lines. These are the highest pump currents for which oscillation on only the grating-selected mode is guaranteed across the entire tuning range. The intersections with the P-I curves of the EC is marked with black dots and their corresponding powers are marked with horizontal lines. For the two different temperatures and the 3 different duty cycles, the output within the pulse is constant with a

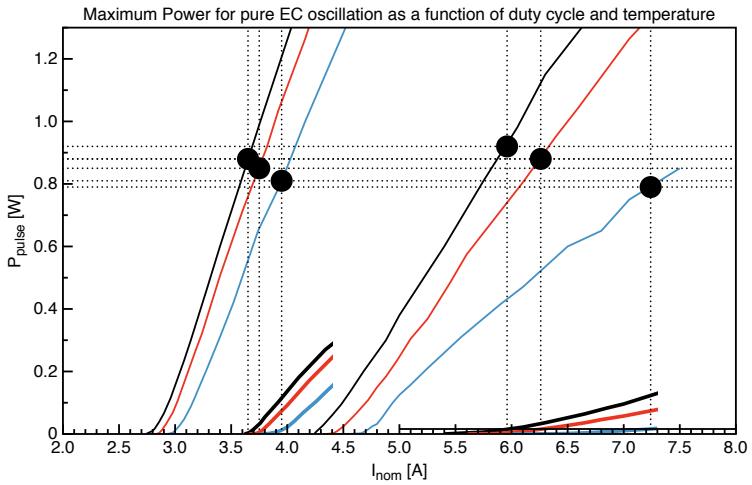


Figure 4.20 *P-I curves from Fig. 4.19 for the grating tuned to the gain maximum along with a section of the FP curves from Fig. 4.17 around the thresholds.*

remarkably small experimental error and has a value of (0.85±0.05) W. This is in very good agreement with the theoretical statement that the maximum, pure EC output power (during a pulse) is independent of temperature. Of course, the time-averaged power *can* be increased through an increase of the duty cycle. This is best done at lower temperatures to increase the lifetime of the QCL, because of the smaller required pump currents.

4.4.2 Power spectra

The power spectrum of an EC is the output power as a function of Littrow wavenumber. Power spectra (pulse power) for the sample HU2-0324 were measured, again, for the two temperatures at the three different duty cycles from the previous subsection. The current values they were measured at are marked in Fig. 4.21 with black dots. These are slightly different to the ones marked in Fig. 4.20, potentially leading to slight mixing with FP modes towards the extremes of the tuning range. Note that even when the ECQCL is pumped slightly above the FP thresholds marked in Fig. 4.20), due to gain clamping, the gain maximum does not reach the FP value except when the grating is tuned to the very extremes of the tuning range. Therefore, there is pure EC operation over nearly the entire tuning range.

The spectra are plotted for a repetition rate of 10 kHz in Fig. 4.22 a), for 87 kHz in b), and for 200 kHz in c). The blue spectra represent the cooled QCL, the red the uncooled QCL. The black horizontal lines mark the corresponding values of the various P-I curves at the marked points in Fig. 4.21, the solid line is for the cooled QCL and the dotted line for the uncooled QCL. This is to illustrate that, in fact, the P-I curves of the EC were taken with the grating tuned to the maximum of the gain spectrum. (The noisy spectrum for the cooled QCL at 10 kHz is due to a poor range setting of the detector in combination with very low average powers at this small duty cycle.)

The first point to notice is that the spectra do not resemble a single clipped Lorentzian as would be expected for a single homogeneously broadened transition, but instead contain more

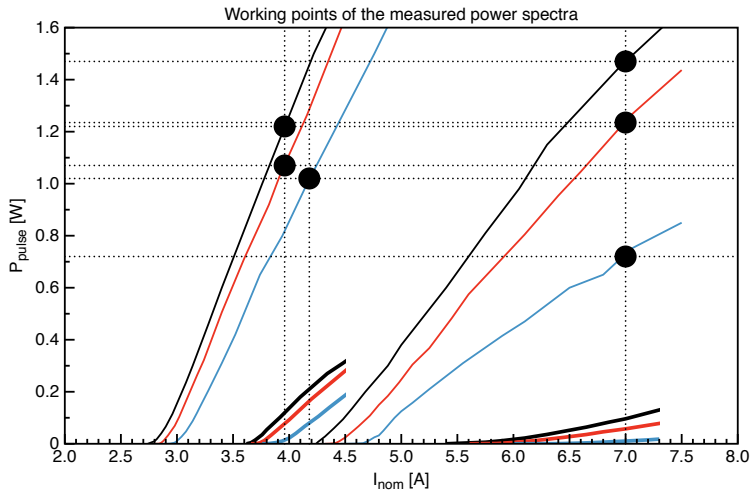


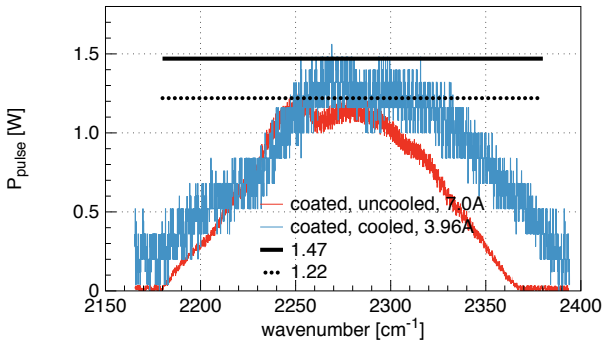
Figure 4.21 *The current values the power spectra were measured at are marked with black dots. These are slightly different to the ones marked in Fig. 4.20.*

structure, that hint to at least 4-5 transitions. This does not seem surprising considering the multiple transitions given the active region in Fig. 4.23 for this particular QCL (denoted by the colored arrows) and the fact that the peaks are located within a span of approximately 5% of their energy. It does however show that a stable ECQCL that can be run at various temperatures is a good tool to investigate the gain of a QCL. This is normally only possible using the electroluminescence spectrum acquired with an FTIR interferometer. That measurement however has a very poor signal to noise ratio, due to the very low overall power levels.

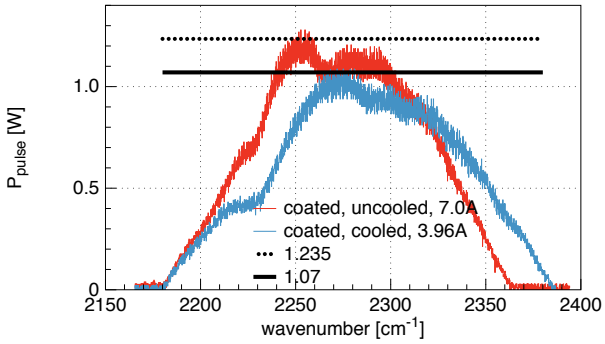
However, the fact that there are multiple transitions also means that the simple expression of the tuning range given in Eq. (1.36) can only be applied with a phenomenological broadening parameter that spans across all the transitions. Also, the tuning range has a smaller temperature dependence than given by Eq. (1.40). This is because the overall tuning range is dominated by the fact that there are multiple transitions whose positions are not strongly temperature-dependent. This way only the temperature-dependences of the lowest-energy and highest-energy transitions contribute to the overall width.

A simulation and discussion similar to Fig. 1.15 in Section 1.5.3 will not be repeated in this chapter, since the parameters used for the simulations there are the effective parameters fitted to the data of the EC-QCL of this section with the correct AR coating of 1.5%.

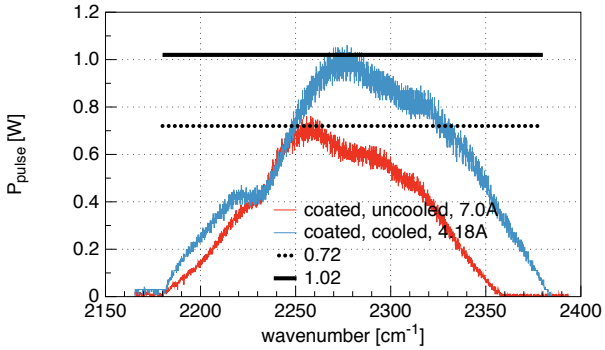
The noise on the spectra (except for the cooled, 10 kHz spectrum) is due to slight intensity fluctuations due to the coupled cavity effect that stems from the small residual reflectivity of the AR coating. The (root-mean-square) intensity fluctuations are approximately 4% of the intensity for both cooled and uncooled EC-QCLs. The housing of the EC is purged with dry nitrogen, thus the CO_2 absorption lines imposed for instance on the spectrum in Fig. 3.7 are not visible here. In fact, it is interesting to compare the uncooled spectrum at 87 kHz in Fig. 4.22 b) with the power spectrum in Fig. 3.7. Figure 3.7 is taken with the same coated QCL at approximately the same duty cycle and pump current. However the tuning range there



(a) *10 kHz*



(b) *87 kHz*



(c) *200 kHz*

Figure 4.22 *Plot of the power spectra at three duty cycles and two temperatures.*

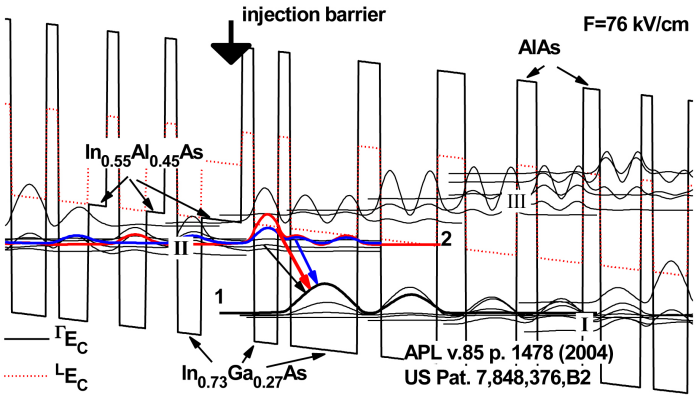


Figure 4.23 *EC max.*

extends only from $(2270 - 2320)\,\mathrm{cm}^{-1}$. This is because of the much smaller feedback due to the reduced collection efficiency of the parabolic mirror compared with the grating. This also leads to a smaller power of the V-ECQCL. The maximum pulse power of Fig. 3.7 is 15 mW, while the maximum pulse power of Fig. 4.22 b) (uncooled) is 1.2 W.

To compare the shape of the power spectra as well as their tuning ranges better, the plots from Fig. 4.22 are normalized to the same height and plotted together in Fig. 4.24. The 10 kHz cooled spectrum is not drawn because the excessive noise would distract the eye. Also, for clarity, the spectra are smoothed by averaging over 10 values in a moving average. As can be seen, the two spectra for the cooled QCL (87 kHz and 200 kHz) overlap perfectly, and the curves for the uncooled QCL (10 kHz and 87 kHz) overlap perfectly.

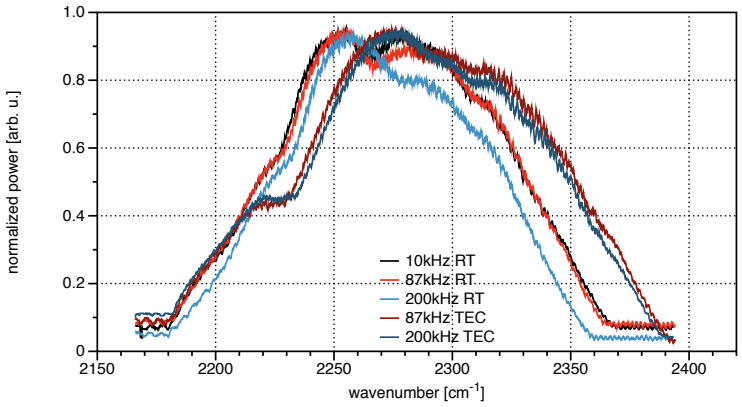


Figure 4.24 *Tuning normalized (to filter out noise, the values have been averaged over the last 10 values (moving average))*

The 200-kHz RT curve (light blue) is slightly narrower, which is expected due to the weaker pumping (relative to the FP threshold current), as can be seen in Fig. 4.21 (this, of course, does not manifest itself in the peak power in this plot, because the graphs are normalized). Due to this fact, it may be best to ignore this spectrum in the further discussion.

As can be seen in Fig. 4.24, the TEC curves are slightly broader, by approximately 15%,

and shifted toward higher energies. This behavior is contrary to what one would expect for a diagonal-transition QCL, where the lower pump currents that accompany operation at lower temperatures also lead to lower voltages. But due to the stark shift, this should result in a red shift for lower temperatures. One might also guess that if there are multiple upper laser levels, then at higher temperatures, a higher-lying laser levels would be more strongly thermally occupied, leading again to a blue shift at higher temperatures. The effect is also contrary to the behavior expected from Eq. (1.40), where $\gamma_{32}(T)$ grows with increasing temperatures. But it has been stated there that if there are multiple transitions, the tuning range and thus its temperature dependence will be dominated by the positions of the transition peaks.

So one explanation for the observed blue shift at lower temperatures here is that either a higher-lying upper laser level has an increased life time, or there is increased coupling to a lower-lying lower laser level. Another possible explanation is that there is a spatial drift in the layer thicknesses and transition energies, and that through slight thermal expansion of the laser mount, the feedback focusses on slightly different areas and thus triggers different transition energies. However, the effect is very small, so it could very well be a simple experimental anomaly.

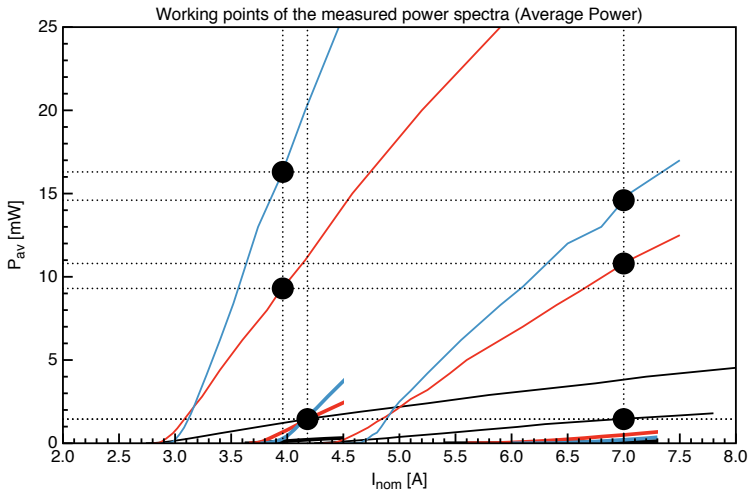


Figure 4.25 *Same data as Fig. (4.21), but for average powers and not pulse powers.*

Finally, Fig. (4.25) shows the same data as Fig. (4.21), but for average powers and not pulse powers. Thus it can be seen, that maximum average power of over 16 mW is reached at a duty cycle of 200 kHz at TEC temperature. Here the EC also has the widest tuning range of approximately $210\,\mathrm{cm}^{-1}$ spanning from $(2180 - 2390)\,\mathrm{cm}^{-1}$, which is nearly 10% of the central frequency.

4.5 First Applications

4.5.1 Measurements of water and CO2 lines

First preliminary spectroscopic absorption measurements have been carried out using the S-ECQCL with several coated QCLs in conjunction with a simple power meter in ambient air (without a gas cell).

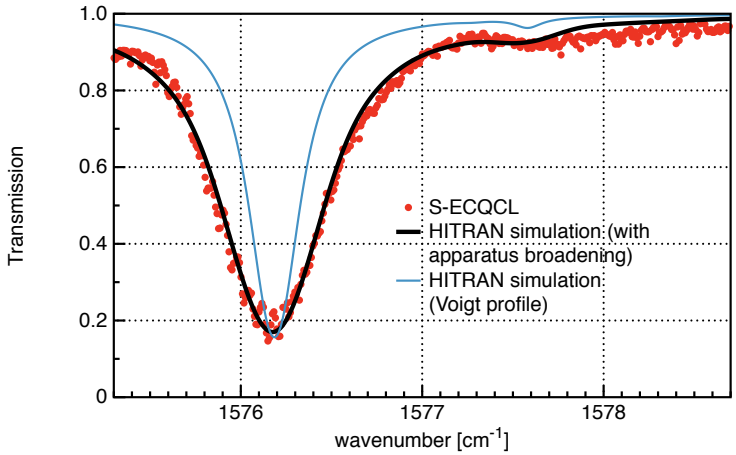


Figure 4.26 *High-resolution scan (red dots) of a water vapor absorption line located at* $1576.2\,\mathrm{cm}^{-1}$ *along with a HITRAN [137] simulation (black line). The Voigt profile of the water vapor absorption is also plotted (blue line).*

Figure 4.26 shows a high-resolution scan (red dots) of a water vapor absorption line located at $1576.2\,\mathrm{cm}^{-1}$ along with a HITRAN [137] simulation (black line). The distance between the Laser head and the bolometer was 15 cm. The spectrum was taken using a QCL that was tunable between $1570\,\mathrm{cm}^{-1}$ and $1620\,\mathrm{cm}^{-1}$. The simulation parameters for the HITRAN spectrum are 1.863% water vapor in ambient air at 296 K.

The shape of the absorption line is determined by a convolution of four contributions. First, there is the Lorentzian line shape of the transition at zero pressure and temperature, whose width is inversely proportional to the lifetime of the excited state. Second, there is the finite pressure that leads to collisions with other molecules, which cause relaxation of the excited state. This broadens the line but preserves the Lorentzian shape, because it effectively reduces the mean lifetime of the excited state. Third, there is temperature broadening due to Doppler-shifted absorption of particles with a relative velocity to the observer. This effect has a Gaussian line shape because of the statistical distribution of the particle velocities. The convolution of these effects results in the so-called *Voigt profile* of the absorption line (at finite pressure and temperature).

Finally, the experimentally observed spectrum is a convolution of the Voigt profile and the *apparatus function*. If the spectrometer uses a collimated laser as the probe beam, the apparatus function is essentially the line shape of the probe beam.

If the linewidth of the apparatus is much smaller than the width of the Voigt profile, the

Voigt profile (and its linewidth) can be observed directly through experiment. If, on the other hand, the linewidth of the Voigt profile is much narrower or of the same order as the apparatus function, the linewidth of the apparatus can be determined by carrying out the convolution of the apparatus function with the simulated Voigt profile and fitting the result to the observed spectrum.

The Voigt profile of the water vapor absorption is also plotted in Fig. 4.26 (blue line), and as can be seen, it is narrower than the linewidth of the apparatus. From the fit of the convolution with the apparatus function, the linewidth of the S-ECQCL in this particular configuration was determined to be $0.35\,\mathrm{cm}^{-1}$.

Figure 4.27 shows a high-resolution scan (green line) of the overlap region of the absorptions of the two most abundant CO_2 isotopologues in the range $(2270 - 2300)\,\mathrm{cm}^{-1}$ along with a HITRAN [137] simulation (black line). The distance between the Laser head and the bolometer was 65 cm. The spectrum was taken using the QCL HU2-0324 discussed in the previous section. The simulation parameters for the HITRAN spectrum are 0.0327% CO_2 in ambient air at 296 K with a relative $C^{13}O_2$ content of 1.106%. The HITRAN curve was fitted with an apparatus resolution of $0.35\,\mathrm{cm}^{-1}$. The deviation of the green from the black spectrum is mainly due to imperfect normalization. Due to the lack of a gas cell, this normalization was done using a background spectrum taken with the bolometer at zero distance from the laser head to avoid CO_2 absorption and then normalized with a constant factor to make up for the beam divergence.

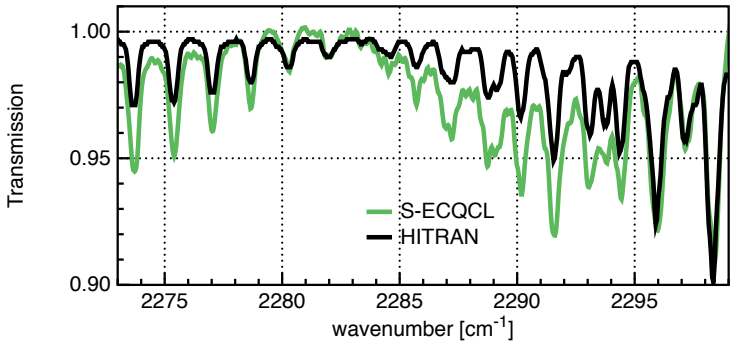


Figure 4.27 *high-resolution scan (green line) of the overlap region of the absorptions of the two most abundant CO_2 isotopologues in the range* $(2270 - 2300)\,\mathrm{cm}^{-1}$ *along with a HITRAN simulation (black line). The peaks to the left of* $2283\,\mathrm{cm}^{-1}$ *are due to absorption from $C^{13}O_2$ and the ones on the right mainly due to $C^{12}O_2$.*

The peaks to the left of $2283\,\mathrm{cm}^{-1}$ are due to absorption from $C^{13}O_2$ and the ones on the right mainly due to $C^{12}O_2$. The abundance of $C^{13}O_2$ is approximately 3.5 ppm and the absorption lines are very clearly visible. With a standard multi pass gas cell with a path length of 50 m, low ppb detectivities are possible. Even higher detectivities could be achieved with high-Q multipass cells.

4.5.2 Measurements of filter transmission

The transmission of the filter discussed in Section 5.4 has also been measured using the S-ECQCL with the QCL HU2-0324. This resulted in Figs. 5.22 and 5.23.

4.5.3 Other Absorption measurements

The S-ECQCL has been dispatched to another research group to perform pump-probe measurements on the dye Coumarin 314 in conjunction with an ultra-violet femtosecond laser (currently not published yet).

5

Passively Alignment-stabilized Filter-tuned External-Cavity QCL (F-ECQCL)

5.1 Introduction

Ever since its first experimental demonstration in 1994 [10], the Quantum Cascade Laser (QCL) has been the main candidate to aspire to similar significance in the mid-infrared (MIR) as the inter-band diode laser has in the visible (VIS) and near infrared (NIR) ranges [16]. While the VIS and NIR provide the most suitable ranges for telecommunication and illumination, the MIR, as discussed, contains both the infrared atmospheric window and the so-called *fingerprint region* of many molecules [2], promising ultra-sensitive yet affordable trace-compound detection in potentially the same package size as a DVD drive or a laser pointer.

On the road to these useful devices, there are challenges to be overcome, though. One stems from the fact that the prototypical application for a QCL involves wavelength tuning, best over a wide range, whereas – with the exception of their scientific use – the bulk of real-world applications of inter-band lasers are fixed-wavelength. While only for the simplest of applications – like detection of single compounds – compact and rugged Distributed Feedback (DFB) QCLs [8] can be used, for most complex tasks – like the analysis of a mixtures – External Cavity (EC) resonators are necessary. These setups tend to be extremely sensitive to mechanical perturbation [145] and are therefore mostly used in a protected laboratory environment. Only few manufacturers produce portable devices by employing ultra-stiff structures that are consequently bulky and costly. And while in the near future, QCLs are expected to drop in price significantly, high-precision opto-mechanics will not.

There is, however, a fundamentally different approach to the most commonly used Littrow and Littman-Metcalf [146] designs, which solves the issue of mechanical sensitivity by using a retroreflector instead of the commonly used diffraction grating [145, 147]. Since retroreflectors

always reflect incoming light back to its source regardless of its direction, this setup can not be misaligned without physically damaging the structure. Figure 5.1 shows a sketch of the alignment tolerance of such a design where the retroreflector is represented by a two-dimensional corner-cube. Therefore, it can be designed in a light-weight fashion without compromising its robustness.

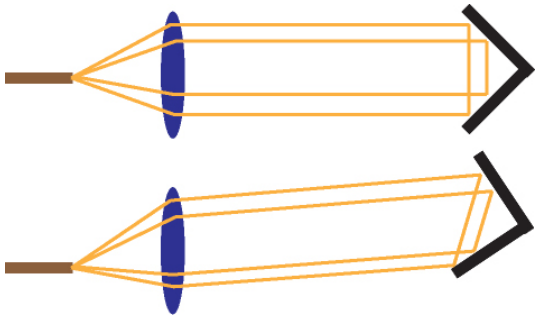


Figure 5.1 *Sketch of the alignment tolerance of a stabilized feedback design where the retroreflector is represented by a two-dimensional corner-cube.*

The wavelength in this case is selected by an interference filter, which has the added advantage that these have much larger angular dispersions than gratings, greatly reducing the spectral fluctuation, while increasing the wavelength accuracy. It also allows for a greater level of miniaturization than grating-tuned setups for many reasons. One, the beam diameter can be nearly arbitrarily small without the filter losing selectivity. Two, due to the great angular dispersion, smaller stepping motors can be used without losing spectral accuracy [145]. Three, filters can be nearly arbitrarily light, making them more fitting as moving parts than bulky gratings. Finally, due to the low cost of volume-produced interference filters and cheaper employable motors, there is a far greater potential for cost reduction.

While grating-tuned setups have become widely established as the state of the art for QCL-tuning since 2002 [22] and set the bar with continuous-wave (CW) emission [27], tuning ranges up to 432 cm^{-1} [6], and mode-hop free fine-tuning [29], filter-tuned designs have only recently manifested themselves and only in the VIS-NIR. In the MIR, "material problems" pose great challenges on producing suitable filters. The following sections, as well as one of our previous publications [193] provide a detailed discussion on these. For this reason, a filter-tuned design has not been demonstrated for MIR sources such as the QCL before this work.

Recently, we have developed a suitable angle-tunable filter with a passband narrow enough to discriminate between laser modes [193]. It has a FWHM of 3.2 cm^{-1} which is similar to the selectivity of a typical 300 rules/mm grating in Littrow configuration for this wavelength. In the following sections we demonstrate its design, fabrication and use as a first Proof of Concept in an alignment-stabilized interference filter-tuned EC-QCL. These results agree very well with a model prediction made earlier [195].

5.2 Design, Simulation, and Fabrication of MIR Bandpass Interference Filters

Thin film interference filters have been a mature technology for at least six decades [148]. Although filters for a wide variety of applications exist off the shelf for the visible spectrum, not many are available for the mid to far infrared [149]. Thus when a mid to far infrared application requires a particular interference filter, it is most probably impossible to be commercially acquired, especially for research applications, i.e. in small production numbers for a realistic cost. With the growing maturity level of modern infrared sources such as the quantum-cascade laser (QCL) [10] and optical parametric oscillators (OPOs), the number of infrared applications is currently exploding [150], requiring a plethora of specialized optics and filters for the mid infrared (MIR), much like for the visible and near-infrared ranges during the advent of the diode laser. Not only is the availability of specialized filters for the MIR quite limited, but fabrication information is as well, since little has been published on these in general lately, resulting in the most informative contemporary resources being product brochures of coating material manufacturers [151]. Additionally, relatively few materials lend themselves to infrared applications in general, due to the reduced transparency in the vicinity of lattice vibrations that tend to be in the MIR [152]. The few materials that do have good optical properties are often either toxic or radioactive [149, 152]. Furthermore, they tend to have very poor mechanical properties, being hygroscopic or prone to cracking, delamination, or softness [152–154]; these properties introduce additional challenges in developing and working with MIR filters.

Bandpass interference filters have not been incorporated into MIR external-cavity lasers such as ECQCLs because, in addition to the necessity for mechanical robustness and optical transparency, the transmission band needs to be narrower than the Fabry-Perot longitudinal laser modes in order to run stably with single-mode emission. For laser stripe lengths of several mm, this means that the transmission band needs to be narrower than a few cm^{-1}.

5.2.1 First considerations

5.2.1.1 Filter type

A filter suitable to select a single mode of a QCL has to have extremely narrow transmission paired with a blocking range at least as wide as the gain of the QCL. Thus it is reasonable to – a-priori – ignore filter designs that pair a low-pass with a high-pass filter to define the transmission range, and instead to focus on designs that are purely blocking in nature with only very thin transmissions where a strict resonance condition is met. The type of design that immediately springs to mind is the Fabry-Perot (FP) cavity.

An FP cavity consists of two plane parallel reflectors facing each other. It has a periodic transmission characteristic that has maximum transmission at the spectral positions where – simplistically speaking, and only in normal incidence – an integral number of half of the

wavelength "fits" between the reflectors. Or – as it is more commonly stated, although not any less simplistic – where the partial waves from the multiple internal reflections and transmissions interfere constructively in the forward direction.

Figure 5.2 (left side) shows a crude ray-optics sketch of an FP cavity in the form of an etalon, where the reflectors are just the polished left and right surfaces of it. A ray is incident on the left that splits up at the first reflector into a ray that is reflected and one that is transmitted. The transmitted ray then hits the second reflector where it, again, is partially transmitted and reflected. The (internally) reflected ray then impinges on the first reflector again, where the process reiterates itself. If the partially transmitted rays, denoted in the sketch as T_1, T_2, and so on, interfere constructively, the overall transmission of the cavity is at a maximum. The typical spectral characteristic of an FP cavity is shown in Fig. 5.2 (right side) for two different finesses.

The expressions given in the following are over-simplified and only valid under very narrow circumstances and only intended to give an intuitive picture for rough estimates. However, the more general analytical discussion is not given in this thesis, since it is more cumbersome without providing much more insight, and is fully incorporated in the most general formalism, the Matrix Method, discussed in Section 2.2, that all the simulations of this chapter are based on. The simulation code is given in Appendix C.2.

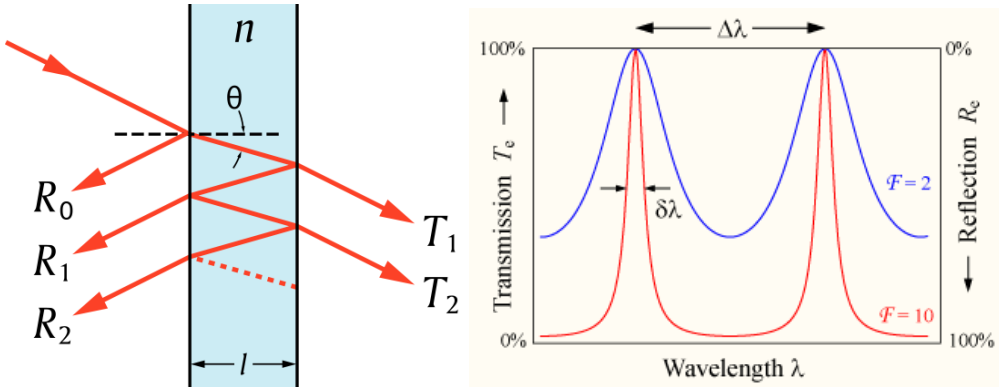


Figure 5.2 *Left: Crude ray-optics sketch of an FP cavity in the form of an etalon, where the reflectors are just the polished left and right surfaces of it. A ray is incident on the left that splits up at the first reflector into a beam that is reflected and one that is transmitted. The transmitted beam then hits the second reflector where it, again, is partially transmitted and reflected. The (internally) reflected beam then impinges on the first reflector again, where the process repeats itself. If the partial beams, denoted in the sketch as T_1, T_2, and so on, interfere constructively, the overall transmission of the cavity is at a maximum. Right: The typical spectral characteristic of an FP cavity for two different finesses. Graphics from [155].*

Since the internal reflections and transmissions at the reflectors interfere constructively in the forward direction for a transmission peak to occur, the transmission peaks are associated with an order, m, which (for normal incidence) are given by:

$$\nu_m = \frac{m}{2nl}, \tag{5.1}$$

where ν_m is the (central) wavenumber of the m-th transmission band, l is the space between

the reflectors and n is the refractive index of the medium between them. This is of course similar behavior as for an FP laser as discussed in Section 1.4.1. Consequently, the range between two transmission bands, the free spectral range (FSR) is given by

$$\mathrm{FSR} = \nu_{m+1} - \nu_m = \frac{1}{2nl}, \tag{5.2}$$

with equidistant spacing in wavenumber or frequency, but not in wavelength. The width of each transmission peak (the full width at half the maximum value, FWHM) can be calculated from a quantity called the finesse to be

$$\mathrm{FWHM} = \frac{\mathrm{FSR}}{F}, \tag{5.3}$$

where the finesse in the range of large reflectivities R is given by

$$F = \frac{\pi\sqrt{R}}{1-R}. \tag{5.4}$$

Another property of filters with FP cavities, that potentially makes it very useful for selecting a laser line is that the transmission peak tunes spectrally with tilt angle of the cavity.

5.2.1.2 Solid Fabry-Perot etalon?

The simplest form of such a cavity is a *solid Fabry-Perot etalon*, a material with polished plane-parallel surfaces, possibly coated with reflective layers. Thus the first question to answer is, whether it is physically possible or feasible from a usability point of view to create or use such an etalon. For this, the projected required specifications of such a filter have to be considered.

A QCL with typical gain may have a gain width of approximately (200-300) cm^{-1}, which in the context of an External Cavity translates nearly identically to the tuning range. Now the filter needs to block all of the light except for the purposely transmitted band, thus the blocking range of the filter has to be at least as great as the tuning range, thus the first condition is

$$\mathrm{FSR} \geq (200 - 300)\,\mathrm{cm}^{-1}. \tag{5.5}$$

Thus considering a design wavelength of $\nu = \nu_m = 2200\ \mathrm{cm}^{-1}$, in order to achieve FSR = $200\,\mathrm{cm}^{-1}$, according to Eqs. (5.1) and (5.2) the order m of the transmission has to be $m \leq 11$. But then, according to Eq. (5.1), the thickness of a solid etalon, made for instance of silicon with a refractive index $n = 3.42$ has to be $l(m = 11) \approx 7.3\,\mu\mathrm{m}$, or thinner for lower orders. But this is clearly too thin of a silicon *membrane* to be mechanically stable when suspended in air. Therefore the etalon has to be stabilized by a substrate it is deposited onto as a structure of thin films.

The filter thus is a thin film interference filter consisting of three parts that need to be considered jointly: The substrate (with an antireflection layer on both sides), reflective layers to define the cavity, and the spacer layer between them.

First, as substrate material, high-quality double-side-polished, infrared-transparent silicon

is available off the shelf. This can be used without any problems. As antireflection (AR) coating, as was discussed in Section 2, yttrium oxide is easily deposited. It turns out, it also has a very useful refractive index for a one-layer AR coating on a silicon substrate, not just on a QCL facet, since the refractive index of silicon (3.4) is very similar to the refractive index of a QCL active region (3.25).

5.2.1.3 Reflective elements of the cavity

Next to consider are the reflective elements of the cavity. To be comparable in transmission bandwidth with a typical Grating in Littrow setups, the FWHM needs to be

$$\mathrm{FWHM} \leq (2-6)\,\mathrm{cm}^{-1}. \tag{5.6}$$

From this and Eq. (5.3) it is clear that it is necessary for the finesse to be $F \gtrapprox 100$ for an 11th order cavity. To achieve the necessary finesse for this bandwidth, the reflectivity according to Eqs. (5.4) would have to be $R \approx 97\%$ (for a cavity used in 11th order). For this purpose, it would be easiest to deposit a thin metallic layer, e.g. gold. So next to consider is whether gold is a good material in this context.

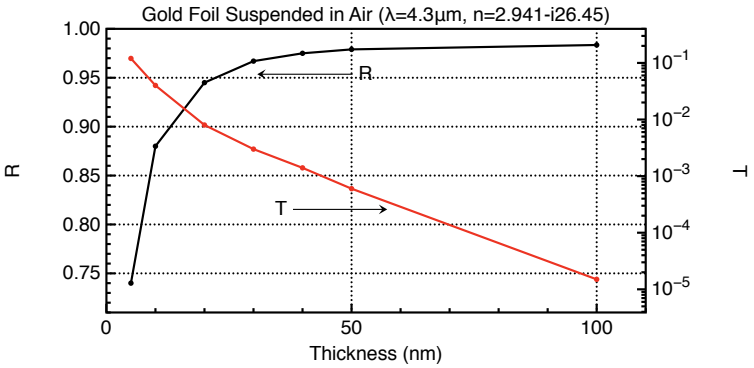


Figure 5.3 *Calculated values of transmission and reflection of a thin layer of gold (in air) as a function of thickness. As the reflection asymptotically reaches its Fresnel value, the transmission exponentially decreases. Thus, the gold foil splits the light into a reflected part and an absorbed part with a negligible transmitted part.*

The refractive index of gold at $\lambda = 4.3\,\mu\mathrm{m}$ is $n \approx 2.941 - i26.45$, where the imaginary part denotes the extinction coefficient. The Fresnel reflectivity of gold (in air) is $R = 98.4\%$, which seems good enough to fit the purposes at first sight. The Fresnel reflectivity, though, is only valid for the boundary to an infinitely thick slab of gold, i.e. $d_{\mathrm{gold}} \gg \lambda$. The actual reflectivity of for instance a 50-nm thick film for $\lambda = 4.3\,\mu\mathrm{m}$ is $R \approx 97.8\%$, which is just about enough to reach the required reflectivity. However, due to the high extinction coefficient of metals that is ultimately responsible for the their high reflectivity, the transmission of a 50-nm thick film of gold is only $T \approx 0.0006$, and the rest, i.e. $1-R-T$ is absorbed, transformed into heat and lost from the process. This part, even for a film as thin as 50 nm, is approximately 40 times greater than the transmitted light. Figure 5.3 shows calculated values of transmission and reflection of a thin layer of gold as a function of thickness. From this plot it is clear, that as the

reflection asymptotically reaches its Fresnel value, the transmission exponentially decreases and thus from comparatively thin values on, the gold foil splits the light into a reflected part and an absorbed part with a negligible transmitted part. Figure 5.4 shows calculated reflectivity (left) and transmissivity (right) curves for a first-order etalon made with very thin, 20 nm gold layers. These plots show how at the resonance the reflectivity decreases sharply (to a value of approximately 60%) and the transmissivity increases. However, the absolute value of the transmissivity is only approximately 2.5% and therefore unacceptable for the targeted application. In conclusion, since all other metals do not behave significantly better, metals have to be ruled out as reflectors for the thin-film etalon.

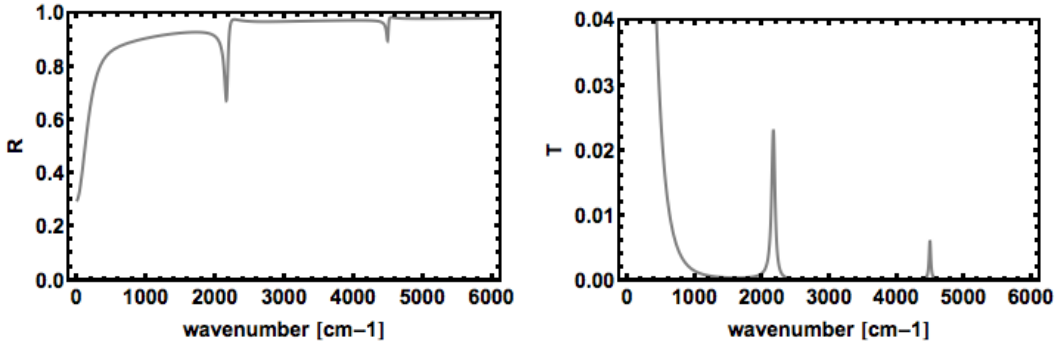


Figure 5.4 *Calculated reflectivity (left) and transmissivity (right) curves for a first-order etalon made with very thin, 20 nm gold layers. At the resonance, the reflectivity decreases sharply (to a value of approximately 70%) and the transmissivity increases. The absolute value of the transmissivity is only approximately 2% and therefore unacceptable for the targeted application*

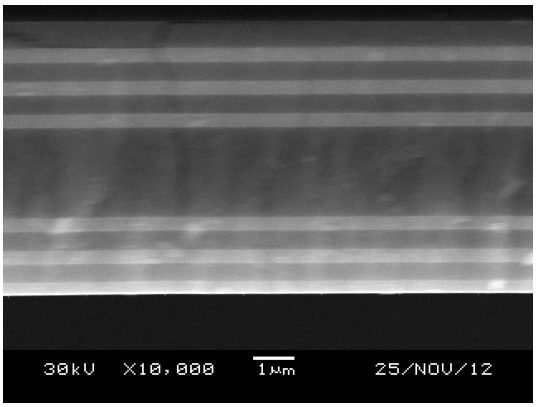


Figure 5.5 *Cross-sectional SEM of a second-order FP cavity made during this work with two DBRs made of three layer pairs each.*

A better way to make mirrors are so-called *Distributed Bragg Reflectors* (DBR), stacks of layers with alternating high and low refractive index and optical thicknesses of a quarter wavelength. It can be shown easily, that the partial internal reflections and transmissions within each pair of layers interferes destructively in the forward direction, reducing the transmission by a constant factor for each successive pair. If both the materials are perfectly transparent, then in principle, arbitrarily high reflectivities can be reached by stacking more and more pairs of layers on top of each other without any absorption. The greater the refractive index

contrast between the two materials, the fewer pairs are required to reach a desired reflectivity. The overall reflectivity is calculated through

$$R_{\mathrm{DBR}} = \left(\frac{n_0 (n_2)^{2\bar{N}} - n_s (n_1)^{2\bar{N}}}{n_0 (n_2)^{2\bar{N}} + n_s (n_1)^{2\bar{N}}} \right)^2, \tag{5.7}$$

where n_s, n_1, n_2, n_0 are the refractive indexes of the substrate, material 1, material 2, and ambient, respectively, and $\bar{N}$ the number of pairs.

Two DBRs facing each other is a single-order FP cavity and higher order cavities can be designed by inserting a spacer between them, whereas the spacer can simply be made of one of the two materials. Figure 5.5 shows a cross-sectional SEM of such a design with two DBRs made of three layer pairs each and a second-order spacer.

5.2.1.4 Spacer of the cavity

Next to consider are the required material properties of the spacer layer, but since the spacer is an integral part of the DBRs, in fact it is made of one of the two materials, the following discussion is valid for the DBR materials as well.

As will become clear later, it is not practical to deposit films thick enough to create higher-order cavities. In case the cavity is used in first order though, where FSR $= \nu_1$, it follows from Eqs. (5.3) and (5.6) it follows that F needs to be $\gtrapprox 1000$. Such high finesses pose very strict conditions on the transparency of the materials in use. To analyze this, it is best to introduce the quality factor or *Q-factor* of an optical cavity used on the m-th transmission band,

$$Q = \frac{F}{\mathrm{FSR}} \nu_m. \tag{5.8}$$

The Q-factor is intimately connected to the finesse, for a first-order cavity, where FSR $= \nu_1$, the Q-factor and the finesse are identical, $Q = F$. But the Q-factor of a high-Q cavity is connected to the power leakage from the cavity upon a round trip through

$$Q = f_m T_{\mathrm{rt}} \frac{2\pi}{x}, \tag{5.9}$$

where $f_m = c\nu_m$ is the optical frequency of the resonating light (of m-th order), $T_{\mathrm{rt}} = \frac{2nl}{c}$ is the time it takes for the light to undergo one round trip, c is the speed of light in vacuum, and x is the fractional power loss during one round trip. But this also means that

$$Q = c\nu_m \frac{2nl}{c} \frac{2\pi}{x} = \frac{\nu_m}{FSR} \frac{2\pi}{x} = m \frac{2\pi}{x}. \tag{5.10}$$

Making use of the fractional round-trip loss x, the power in the cavity at time t is

$$P(t) = P(0)(1-x)^{t/T_{rt}} \tag{5.11}$$

or in terms of the number $\tilde{N}$ of oscillations undergone

$$P(\tilde{N}) = P(0)(1-x)^{\tilde{N}}. \tag{5.12}$$

For low-order cavities, if the Q-factor is large, the round-trip loss is small, as is clear from Eq. (5.10). In that case, Eq. (5.12) becomes

$$P(\tilde{N}) = P(0)(1-x)^N \approx P(0)e^{-x\tilde{N}} \tag{5.13}$$

and the number of round trips it takes for the power inside the cavity to decay to $1/e$ of its original value is then consequently

$$\hat{N} = \frac{1}{x}. \tag{5.14}$$

In this case, however, the Q-factor in Eq. (5.10) becomes

$$Q = 2\pi m\hat{N}. \tag{5.15}$$

Thus the effective distance $\hat{l}$ the light travels inside the cavity (using Eq. (5.8) and $\frac{\nu_m}{\mathrm{FSR}} = m$) is

$$\hat{l} = 2l\hat{N} = \frac{2lmF}{2\pi m} = l\frac{F}{\pi} \tag{5.16}$$

Thus the effective path length through a cavity is not the width of the cavity, l, but is multiplied by a factor of $\frac{F}{\pi}$. Thus for a first order cavity of $F = Q = 1000$, as required for the purpose of laser mode selection, the cavity material must be *highly* transparent for the wavelength range in question to have an appreciable amount of light left when traversing the cavity. In fact, to have a sufficient contrast between blocked and transmitted ranges, it would be desirable if the filter had a transmission T of at least

$$T \gtrapprox (60-80)\%. \tag{5.17}$$

How does this translate to the required transparency of the materials? A plane light wave passing through a medium with a complex index of refraction $\hat{n} = n - ik$ propagates according to the relation

$$\vec{E}(z) = \vec{E}(0)e^{-i(\hat{k}z-\omega t)}, \tag{5.18}$$

where $\vec{E}(0)$ is the electric field vector at $z = 0$, $\hat{k} = \frac{2\pi}{\lambda_0}(n - ik)$ is the complex magnitude of the wave vector, λ_0 is the wavelength in vacuum. At an effective distance $z_0 = \frac{\lambda_0}{2n}\frac{F}{\pi}$, which is the effective optical thickness of a first-order cavity, the intensity of the light is supposed to decay to no less than 60%. This means

$$\left(e^{-\frac{2\pi}{\lambda_0}z_0}\right)^2 = e^{-\frac{4\pi}{\lambda_0}\frac{\lambda_0}{2n}\frac{F}{\pi}} \geq 0.6. \tag{5.19}$$

If the spacer material has an index of refraction of, for instance, $n = 1.5$, this means that the extinction coefficient k has to be $k \lessapprox 4 \times 10^{-4}$, which is very small.

It is important to note here that this is just a relatively rough estimate, since strictly speaking the factor F/π was determined as the effective extension of the cavity on the base

that the bound light is to decay to $1/e$, but the same value is used to determine the required transparency of the material for a decay to 60%. But the error is quite small. For a more accurate value, the full structure is best simulated using the Matrix Model.

The next section discusses the problem of finding suitable materials for infrared DBRs to be used in high-Q cavities, i.e. materials that are transparent enough to be useful in the sense discussed so far.

5.2.2 Choosing the right materials

Appendix A provides the framework to find the best materials possible for infrared interference filters out of all possible compounds of the elements of the periodic table. The line of thought for the selection is presented in this section.

For reasons that became clear in Section 5.2.1.4 and 5.2.1.3, the materials in use for high-Q interference filters need to have a very high level of transparency and to reduce the number of layers required to reach a certain reflectivity and thus finesse, a high refractive index contrast has to be chosen between the two alternating materials. These are consequently called the H-material and the L-material, for high and low index, respectively.

For high transparency in the infrared, it is first of all necessary to have very few free carriers. This obviously rules out all metals as potential materials, and the use of semiconductors strongly depends on the exact wavelength range in question, to determine whether the (Drude) free carrier absorption has tailed off to an acceptable level. Even for the smallest of band gaps, obviously bound by the threshold to interband transitions, the intrinsic absorption will be negligible, but during deposition, close care has to be taken to ensure purity, since most dopants in the form of contaminations will increase the number of free carriers available, and thus increase absorption in the infrared. More importantly though, contaminations, especially water and carbonates, even if not thermally ionized, tend to often have their absorption bands in the mid-infrared, as discussed in Appendix A.4.5.

It is clear from Appendix A.4.6 and the Kramers-Kronig relations, Appendix A.4.1 that highly covalent materials with small band gaps have large refractive indices in the infrared range. The use of monoatomic semiconductors has another great advantage, which is the absence of phononic absorption bands, which also tend to be in the infrared, thus reducing transparency there. A very good candidate for an H-material is consequently germanium with a refractive index of approximately 4.2 in the range between 2 and 12 μm and no absorption if contaminations are kept low. The absorption coefficient within the range of interest for Ge is plotted in Fig. 5.6 (left side). As can be seen, in the range between 2 and 12 μm the absorption is less than 0.04 cm^{-1}. The refractive index is plotted in Fig. 5.6 (right side) as measured for a layer deposited in this work using the SENDIRA Spectroscopic MIR Ellipsometer at Sentech Instruments, see Appendix B. The refractive index changes only very slightly in this range and absorption is lower than the detection limit. Due to interband transitions, materials with a low enough band-gap to achieve a refractive index this high are not transparent in the visible range. Thus, in the visible, the generally attainable refractive

indexes are limited by the value of diamond, which is 2.4. Germanium is extensively used as an H-index material in the infrared due to this advantageous property and the fact that it has very good mechanical and evaporative properties, discussed in a later section.

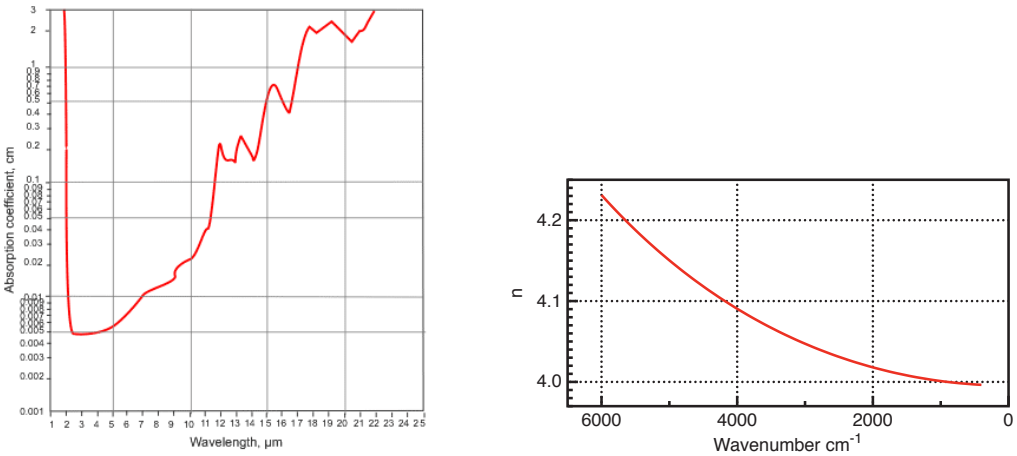


Figure 5.6 *Left side: The absorption coefficient within the range of interest for Ge. In the range between 2 and 12 µm the absorption is less than 0.04 cm^{-1} (Data from [156]). Right side: The refractive index as measured for a layer deposited in this work using the SENDIRA Spectroscopic MIR Ellipsometer at Sentech Instruments. The refractive index changes only very slightly in this range and absorption is lower than the detection limit.*

A much greater challenge is finding a suitable candidate as an L-material. This is the inhibiting factor, which to this day makes high-performance infrared interference filters very few and far-between. Especially angle-tunable filters with the performance achieved in this work have to the best of our knowledge not been produced.

The challenges will become clear in the following. In order to attain a low refractive index in the infrared, materials with a high band gap and low electronic polarizability have to be used. Low electronic polarizability requires highly ionic materials with small ions. But ionic materials have phononic absorption bands that tend to lie in the infrared, thereby diminishing the transparency close to them. So for transparency it is important to choose ionic materials with high masses so the phonon bands are far away in the infrared. Thus there is a tradeoff between transparency and low refractivity. Also the softer the material is, the smaller the inter-ionic potential, and thus the lower the phonon energy. But soft materials have the drawback of mechanical instability.

A closer look at the strongly ionic materials of the periodic table reveals another great challenge of a different kind. The alkali halides tend to be highly hygroscopic and many are highly toxic. Absorption of water from ambient air diminishes the lifetime of the filter and even the smallest quantities of water strongly absorb light in the infrared. So the hygroscopic materials like sodium chloride cannot even withstand the handling prior to deposition without including enough water to destroy the transparency properties. Also, the lower the binding energy of the ions, the more they dissociate during evaporation, producing sub-stoichiometric layers, that have bad transparency. The dissociated ions also tend to corrode the vacuum pumps and tubes and tend to be very toxic. The toxicity obviously puts constraints on the

deposition methods and machinery usable for deposition for reasons of work safety. The best compromise in terms of hygroscopicity and stability are the fluorides. Coincidentally, fluoride is also the smallest ion, which is very advantageous in terms of electronic polarizability. But since it is also the lightest ion, a heavy cation has to be paired with it to achieve somewhat low phononic energies necessary for good transparency.

However, the heavy metal fluorides tend to have very bad deposition characteristics when evaporated, despite not dissociating. That is, they tend to form highly strained layers that are prone to cracking, especially at the very thick layers required for infrared coatings due to the greater wavelength compared with the visible range. Also, a problem of ionic materials in general is that they do not adhere well to covalent materials such as germanium, and therefore tend to delaminate, especially when under great stress.

The stress is due to the fact that when evaporated, whether thermally or with an electron beam, the molecules have very little kinetic energy paired with a high sticking coefficient, thus they tend to stay exactly where they first hit, i.e. the evaporation of ions is termed a "line of sight" process. But this leads to not densely packed layers but columnar structures with voids between them[157]. These voids have two disadvantages. The first is they tend to collapse due to cohesion, forming the tensile stress. The second is that even when collapsed, the layers are very porous and therefore absorb water as soon as they are exposed to ambient humidity.

The optical properties of YF_3 are plotted in Fig. 5.7, as deposited by us and measured using the SENDIRA Spectroscopic MIR Ellipsometer at Sentech Instruments on bare 500 nm single-layer samples. The absorption in YF_3 from approximately 2000-3700 cm^{-1} stems from the absorbed water. This behavior is not found in Ge of Fig. 5.6.

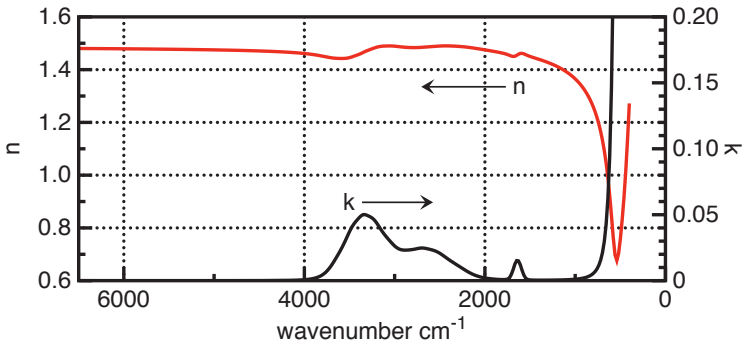


Figure 5.7 *Optical MIR Properties of YF_3 measured using the SENDIRA Spectroscopic MIR Ellipsometer at Sentech Instruments.*

One way of reducing the porosity along with the stress is by heating the substrate to high temperatures, thereby giving the adatoms the necessary kinetic energy. But this leads to polycrystalline layers that have very rough surfaces leading to interface scattering in addition to scattering between the microcrystals within the layer. Also, due to the different thermal expansion coefficients of the layers and the substrate, there is once again stress build-up when the sample cools to ambient temperatures. Thus the trade-off is between porous stressed material with absorbed water and stressed polycrystalline material with high levels of interface

roughness.

Since the mechanical challenges are so profound, due to the lack of simple alternatives, infrared coatings for the longest of time have used thorium fluoride as the L-material, which has comparatively good optical as well as mechanical properties. But thorium is a radioactive α-emitter, making it one of the most toxic materials possible and its use has since been banned for optical coatings.

For this reason, the challenge remained to find a feasible compromise in L-materials, and it turns out that the best choices for transparent materials in the 3-12 μm range are yttrium fluoride, ytterbium fluoride and cerium fluoride, which are all essentially non-toxic and rather stable toward dissociation upon evaporation. Therefore the remaining challenge of producing a high-Q infrared interference filter lied in optimizing the evaporation parameters to reduce the overall stress and water absorption to such an extent as to make the desired filters possible as well as find a suitable material to act as a thin adhesion layer to eliminate the delamination of the layers from germanium.

5.2.3 Development of the fabrication process

As discussed in Section 5.2.2, deposition of fluorides in very transparent layers thick enough for the mid-infrared wavelength range is not straight forward, due to either dissociation when using a high kinetic energy process such as sputtering, or internal stress and delamination when using a low kinetic energy process such as evaporation. These challenges required a lengthy optimization process outlined in this section.

The choice was made to avoid dissociation by evaporating the material gently using an electron beam, and to work instead on decreasing internal stress and delamination. The deposition system in use is a Lesker PVD 75 E-beam evaporator running CWare Software. Reactive gases or plasma assist are not applied.

From the very first test runs, we chose to concentrate on optimizing yttrium fluoride as the L-material. The yttrium fluoride granules are filled into Fabmate crucible liners and loaded into the chamber hearths. 400-μm-thick double-side optically polished silicon substrates are also loaded into the chamber. It is then pumped down to 5×10^{-5} torr.

The first trial layers of yttrium fluoride came out of the deposition chamber in the way illustrated in Fig. 5.8. The photograph shows the sample holder and the $10 \times 20\,\mathrm{mm}^2$ samples with 500 nm thick layers that exhibit very little adhesion, but instead, lots of internal stress.

The first step is to ensure better evaporation behavior without spitting. For this, the yttrium fluoride granules are manually outgassed and pre-melted using the electron beam with a shutter protecting the substrate. The acceleration voltage is 10 kV. Because of the low thermal conductivity of YF_3, the granules are heated with a beam pattern evenly spread out over the entire crucible area, resulting in a melt that covers the entire crucible area at a current of approximately 20 mA. There is a thick layer of molten material, but the material doesn't melt all the way to the bottom of the crucible.

Stress in the layers was combatted by iteratively elevating the substrate to moderate tem-

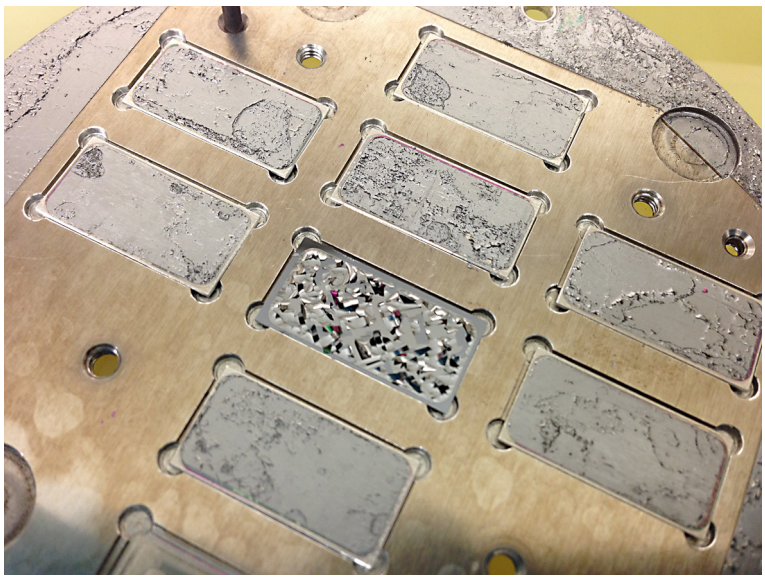


Figure 5.8 *Photograph of the sample holder with* $10 \times 20\,\mathrm{mm}^2$ *samples with unoptimized 500 nm thick layers of* YF_3 *that exhibit very little adhesion, but instead, lots of internal stress*

perature, delamination was reduced by introducing a a very thin, 5-nm adhesion layer of Y_2O_3.

A simple stress test was developed to test all the structures for durability in a reproducible way. For this test, after deposition, the structures are exposed to ambient air for one hour, allowing them to potentially absorb humidity. Afterwards, they are quickly heated to 190°C in a dry nitrogen environment with a ramp time of 5 minutes, held at this temperature for 5 minutes, then ramped down to room temperature (RT) over 5 minutes. The thermal cycling combined with the expansion of the absorbed water often lead to cracking and delamination as is examined and compared using a visible light microscope at 100x magnification.

Figure 5.9 shows some of the first 1000 nm single-layer structures of YF_3 after the stress test under the microscope. Fig. 5.9 a) shows a layer deposited at RT. Clearly visible are large cracks and what seem to be bubbles that we attribute to delamination. Fig. 5.9 b) shows a layer of the same thickness deposited at a substrate temperature of 200°C. The cracks and bubbles are fewer and further apart, indicating generally lower levels of internal stress. Fig. 5.9 c) shows a layer of the same thickness deposited at a substrate temperature of 200°C with the use of the 5-nm adhesion layer. The delamination bubbles are completely absent and the cracks have grown finer. However, thicker layers at this point have still flaked off under the stress test. Figure 5.10 a) shows a violently flaking 2000 nm layer of YF_3 (with adhesion layer) after the stress test.

The next step to combat flaking is by thoroughly preparing the surface using the procedure listed in Table 5.1. Through the hydrofluoric acid bath and subsequent rinsing and drying, the dangling bonds of the silicon surface are temporarily pacified with hydrogen. Therefore the substrates are immediately placed in the vacuum chamber to avoid oxidation. Figure 5.10 b) shows another 2000 nm layer of YF_3 (with adhesion layer) deposited onto a substrate prepared this way after the stress test. Obviously, flaking has seized completely and is now

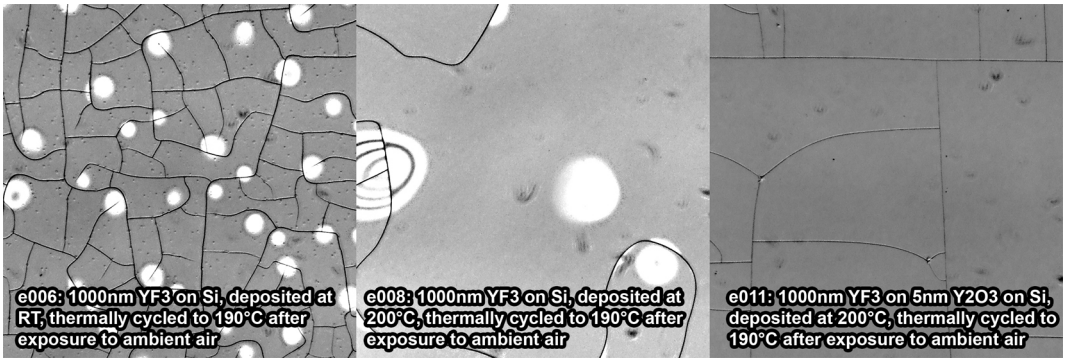


Figure 5.9 *Some of the first 1000 nm single-layer structures of YF_3 after the stress test under the microscope. a) layer deposited at a RT. b) layer of the same thickness deposited at a substrate temperature of 200° C. c) layer of the same thickness deposited at a substrate temperature of 200° C with the use of the 5-nm adhesion layer.*

smooth as seen with visible light. To our best knowledge, none of the literature on infrared coatings has demonstrated smooth and stable YF_3 of this thickness at this point in time.

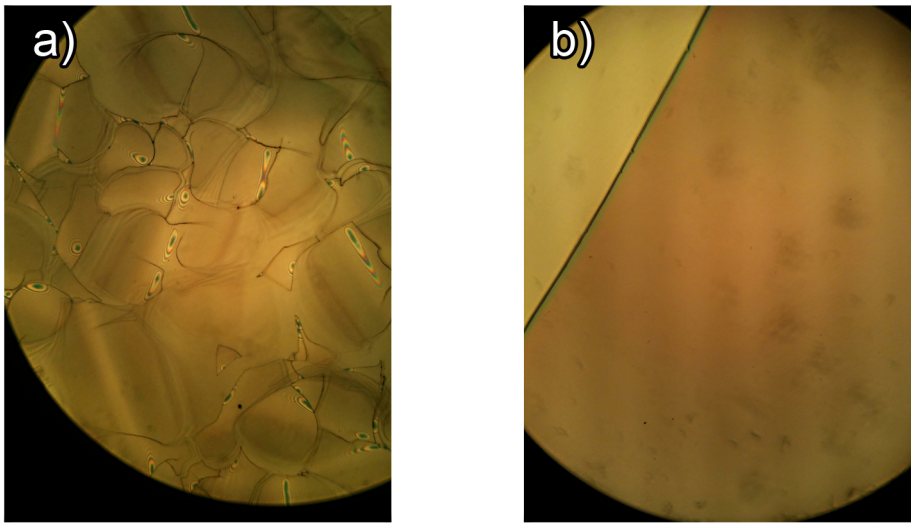


Figure 5.10 *a) Flaking 2000 nm layer of YF_3 (with adhesion layer) after the stress test. b) 2000 nm layer of YF_3 (with adhesion layer) deposited onto a substrate prepared according to the procedure in Table 5.1 after the stress test.*

Next, Ge is prepared by focussing the beam to a small spot in the center of the crucible and manually ramping the current up to 50–100 mA, until the entire content melts and the oxide layer on top breaks open. The process takes approximately 2–3 minutes. Ge melts in bulk like a metal.

Fig. 5.11 shows a cross-sectional scanning electron micrograph (SEM) of the first double-pair test structure grown with alternating layers of nominally 261 nm Ge and 596 nm YF_3 with 5nm Y_2O_3 at every interface at two different magnifications. The samples are smooth to the eye and under the microscope, even after the stress test. As can be seen, the process proved successful so far. The Ge layers appear bright and the YF_3 layers darker. The Y_2O_3 layers are not visible at this magnification.

Table 5.1 *Surface preparation steps.*

No.	Step
1	Rinse with ethanol
2	Blow dry with dry N_2
3	Plasma oxidation up to 15 min
4	HF bath
5	Rinse with deionized water
6	Blow dry with dry N_2
7	Immediately place in deposition chamber
8	Evacuate chamber (while surface is H-passified)
9	Heat in vacuum for deposition

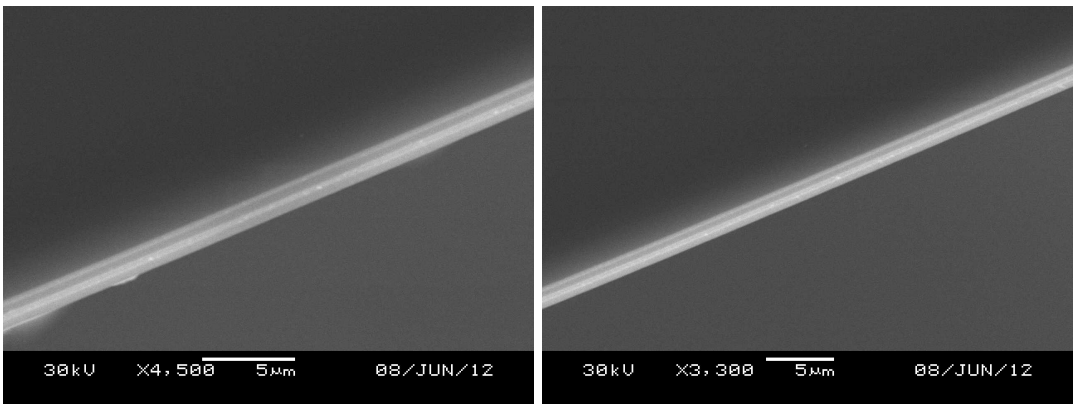


Figure 5.11 *Cross-sectional scanning electron micrograph (SEM) of the first double-pair test structure grown with alternating layers of 261 nm Ge and 596 nm YF_3 with 5 nm Y_2O_3 at every interface at two different magnifications.*

Fig. 5.12 shows a cross-sectional scanning electron micrograph (SEM) of the first tripple-pair test structure grown with alternating layers of nominally 261 nm Ge and 596 nm YF_3 with 5 nm Y_2O_3 at every interface at two different magnifications. The samples are smooth to the eye and under the microscope, even after the stress test. The process is still successful.

To grow the full structure, including the Y_2O_3 anti-reflection coating underneath, a more elaborate Y_2O_3 granule preparation had to be developed. The problem with Y_2O_3 when deposited at larger thicknesses is that the material in the crucible is prone to hole burning, i.e. the formation of small vertical tunnels, with the danger of burning all the way down into the bottom of the crucible liner and crucible, thereby damaging the entire deposition system. This is because Y_2O_3 has a very low thermal conductivity and thus melts in an extremely thin layer. This layer is barely visible and the process of evaporating Y_2O_3 with an electron beam is reminiscent of sublimation. The evaporation rate, however, increases with increasing temperature. Thus every small "dimple" between granules can develop into a quickly deepening tunnel through a runaway effect, since inside the hole, thermal radiation can not escape as well as at a horizontal surface, thus it is much hotter than surrounding areas, causing material to evaporate much faster out of the hole. The solution to this problem

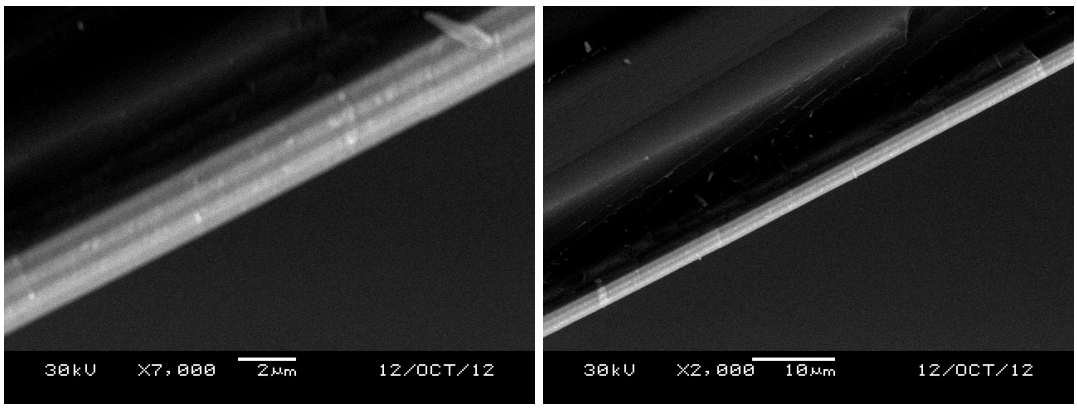


Figure 5.12 *Cross-sectional SEM of the first triple-pair test structure grown with alternating layers of 261 nm Ge and 596 nm YF_3 with 5 nm Y_2O_3 at every interface at two different magnifications.*

takes two steps. First, the crucible content is prepared by focussing the beam to a small spot and ramping up the current carefully to a very high value of 100 mA, while manually sweeping the area of the crucible opening. The beam then locally melts the Y_2O_3 and the granules collapse. The entire area of the crucible is swiped repeatedly to form a dense, smooth surface of bulk, glass-like Y_2O_3 without any dimples. This process is relatively cumbersome and care has to be taken not to use up too much of the material before the actual deposition. Second, the evaporation current is raised to a very high value to make the local melt more pronounced and to allow it to potentially fill any developing holes during deposition.

Fig. 5.13 (left side) shows a cross-sectional SEM of the full filter structure using two DBRs with two layer pairs each and a second-order cavity and the Y_2O_3 AR coating underneath. Fig. 5.13 (right side) shows the full filter structure with three layer pairs. The substrate is toward the top of the image. The first (darker) layer is the Y_2O_3 AR coating and the layer pairs start from top to bottom with Ge. The layers are still smooth and have parallel interfaces. At this stage we began examining the actual transmission characteristic using a Fourier Transform Infrared Interferometer (FTIR). However, both the structures of Fig. 5.13 showed no transmission at all when measured using an FTIR.

To solve this problem, the actual layer thicknesses were from this point on constantly measured and remeasured using the SEM and the thicknesses were iteratively improved over a series of deposition runs, thereby calibrating the software of the deposition system. Several steps (Samples e018, e020, e021) of this process are shown in Fig. 5.14. All samples were measured for transmission as well.

Sample e018 showed no transmission at all in the FTIR and therefore it is not plotted. The transmission spectra of Samples e020 and e021 are plotted in Fig. 5.15. To avoid noise, the transmission spectra (red) are not normalized to the intensity profile of the source, but instead the intensity spectrum of the FTIR source is plotted separately (black). The expected (normalized) reflection and transmission spectra as simulated for these structure are given in Fig. 5.16. The spectra plotted in Fig. 5.16 are for a first-order cavity, two double-layer design. At this magnification, the calculated spectra for structures using two or three layer pairs and for first-order or second-order cavities are barely distinguishable. To distinguish

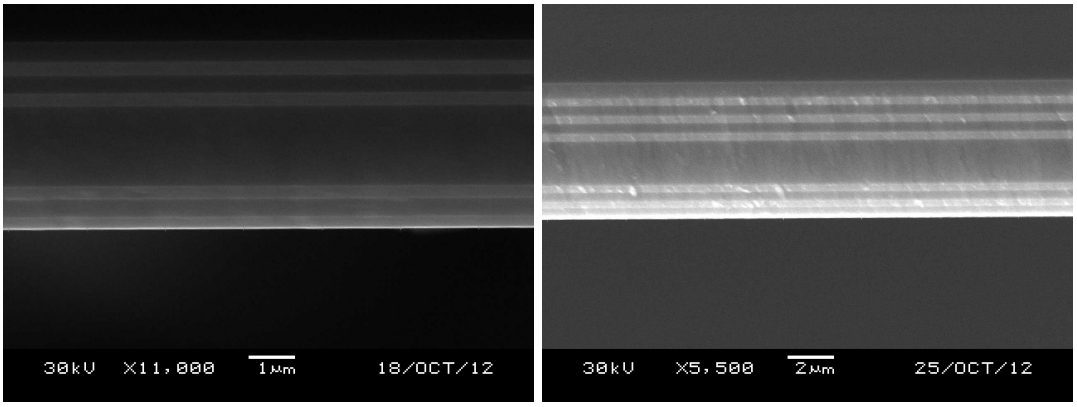


Figure 5.13 *Left side: Cross-sectional SEM of the full filter structure using two DBRs with two layer pairs each and a second-order cavity and the Y_2O_3 AR coating underneath. Right side: Cross-sectional SEM of the full filter structure using two DBRs with three layer pairs each and a second-order cavity and the Y_2O_3 AR coating underneath.*

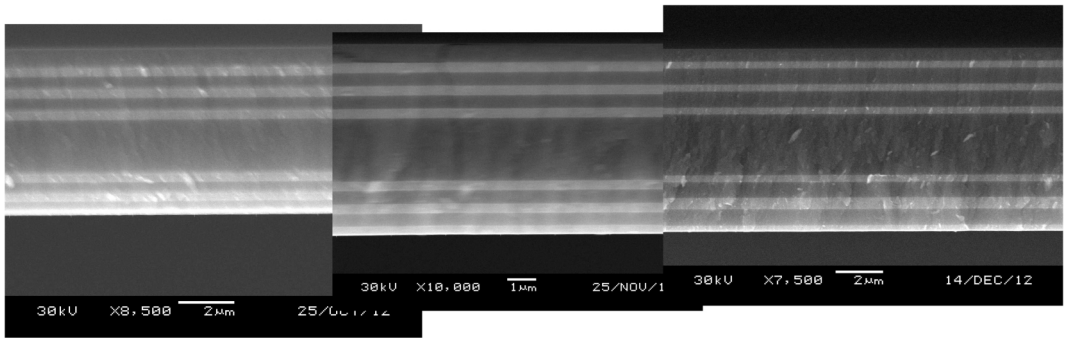


Figure 5.14 *Series of deposition runs with iteratively improved layer thicknesses. Left side: Sample e018. Center: Sample e020. Right side: Sample e021; Note the erroneously thin and thick Ge layers (first and second bright layers from the bottom). The substrate is toward the top of the image. The first (darker) layer is the Y_2O_3 AR coating and the layer pairs start from top to bottom with Ge.*

them, the central peak needs to be strongly magnified. From Fig. 5.16 it can be seen that the stop band should extend from approximately (1200 – 3200) cm^{-1} and the central transmission peak is designed to be at 2365 cm^{-1}.

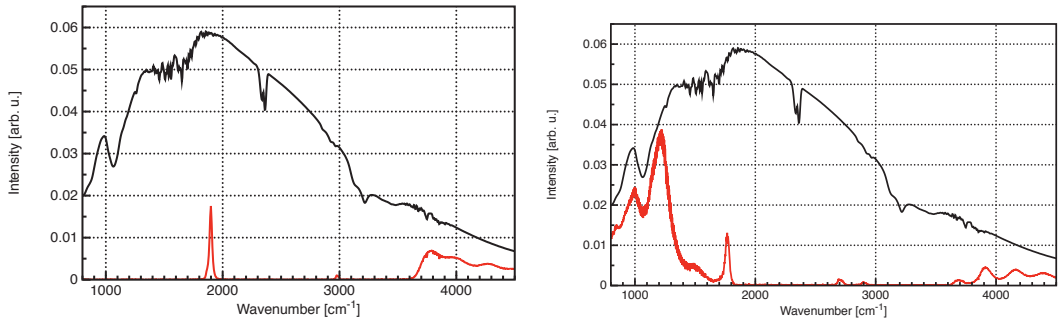


Figure 5.15 *Left side: Transmission spectrum of Sample e020. Right side: Transmission spectrum of the Sample e021. (Note the two small peaks at 2200^{-1} and 2400^{-1})*

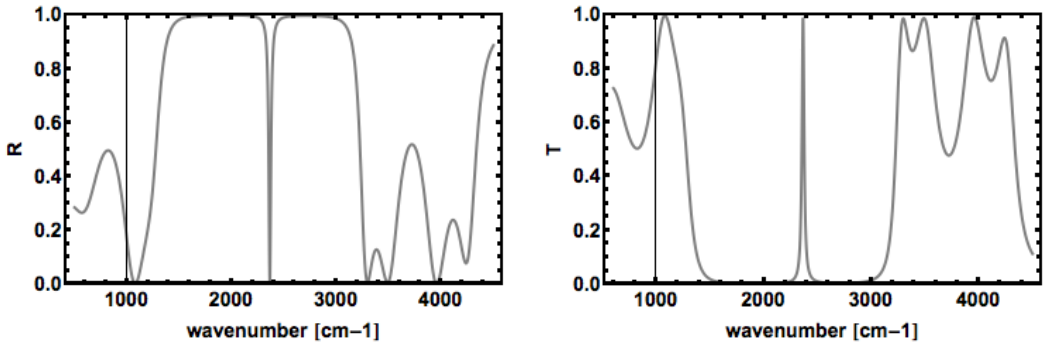


Figure 5.16 *Expected reflection (left) and transmission (right) behavior as simulated for the structure in Fig. 5.13 (left side).*

Sample e020 in 5.15 has a wrong stop band when compared to Fig. 5.16, and the transmission peak located at 1900 cm^{-1} is not the central peak of the structure, but a sideband. The transmission line that was aimed at is barely visible at 3000 cm^{-1}, far from the designed position. From the SEM, it is clear that the Ge layers are much too thick. For Sample e021, the stop band lies much closer to the targeted region, the sideband is at 1750 cm^{-1} is now more clearly recognizable as a sideband. The targeted transmission is visible as two very small peaks at 2700 cm^{-1} and 2900 cm^{-1}. This is in agreement with the SEM in Fig. 5.14 (right), since there is an erroneously thin and an erroneously thick Ge layer (first and second bright layers from the bottom).

To improve on these results, thorough debugging of the entire deposition process had to be undertaken. First, to improve on the precision of the layer thicknesses throughout the entire structure, the microbalance crystal has to be exchanged during a coating run. This is because, due to the large amount of material that needs to be deposited, the crystal reaches the end of its lifetime, thereby losing precision. This can be seen from all the previously shown SEMs, in that the later layers are all thicker or thinner than the earlier layers, which tend to be more homogeneous in thickness.

Now to minimize the amount of material deposited onto each microbalance, it would be best to exchange it in the middle of the coating run. But, since, unfortunately, the vacuum needs to be broken to exchange the microbalance, contaminations are unavoidable at the point in the structure where the chamber was opened. But to reduce the impact of these contaminations on the performance of the structure, they should be at a position where the light is less intense. But this means as far away as possible from the center of the structure. So there is a tradeoff. We found that the best time to open the chamber is in the middle of the second layer pair in a three-pair structure. This is because then the germanium layer is the exposed layer, but Ge is much less porous and prone to incorporation of ambient humidity, thus protecting the YF_3 layers. The first layer to deposit, once the chamber is closed again, is a fresh adhesion layer.

The next step was to simplify the design from a second-order to a first-order cavity, since the layer thickness of the spacer layer is supposed to be an integral multiple of the other layers of the same kind, but the absolute thickness error accumulates with deposition time. Also, at first, the design was simplified from a three double-layer to a two double-layer design to minimize the room for error.

The next step was to increase cleanliness within the chamber by reducing the residual pressure. For this the pumping time after each chamber venting was increased from 2 hours to overnight, and one crucible filled with titanium was also loaded into the chamber and was evaporated (substrate protected by the shutter) to act as a getter material to further reduce chamber pressure. The chamber pressure was reduced by two orders of magnitude this way to 5×10^{-7} torr.

The next step was to decrease the porosity further by elevating the substrate temperature to 300°C. But to avoid cracking upon cooling, the temperature had to be lowered very slowly in a cycling manner, i.e. cool by 100 K slowly, then heat by 50 K, then cool by 100 K and so on.

These final steps increased the processing time of a single deposition run to three days. This could easily be avoided, however, by equipping the deposition system with a second microbalance that is covered by a shutter during the first half of the deposition run.

The final deposition parameters can be found in Table 5.2.

Table 5.2 *Deposition Parameters.*

Material	Energy	Beam Shape	Rate	Approximate Current
Ge	10 keV	Spot	$5\mathring{A}$/s	90-100 mA
YF_3	10 keV	Spread Out	$15\mathring{A}$/s	13-20 mA
Y_2O_3	10 keV	Spread Out	$6\mathring{A}$/s	70 mA

The first sample we grew after performing all of the above optimization steps, was a first-order cavity, two double-layer structure shown in Fig. 5.17. The SEM shows very smooth and parallel layers, whose thicknesses are homogeneous throughout. The corresponding FTIR

transmission spectrum (this time normalized to the intensity profile of the source) shows what turns out to be a beautiful copy of the simulated spectrum from Fig. 5.16. The FWHM of the central transmission peak is 29 cm^{-1}, which can be seen from the inset that is a zoom-in on the peak. The FWHM is less than 3% of its central wavelength, which also corresponds very well to the simulation for a two double-layer design. The overall height of the peak is approximately 70%.

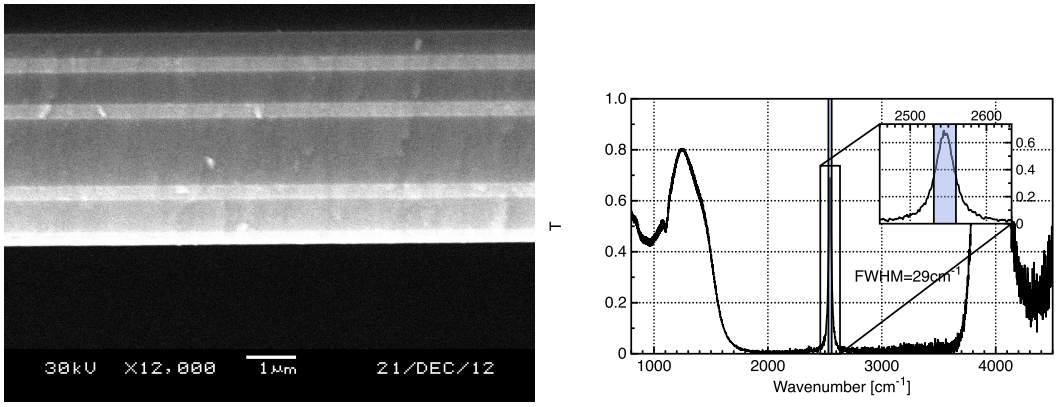


Figure 5.17 *SEM and source-normalized transmission spectrum of the first structure (first-order cavity, two double-layers) grown after the debug process. Note the similarity between this measured spectrum and the calculated spectrum of Fig. 5.16. Inset: Zoom-in on the central transmission peak.*

After the great success of the previous coating run, the design was adjusted back to three layer-pairs. The result of this run is shown in Fig. 5.18. The width of the central peak decreased by a factor of three to 9 cm^{-1}, but the overall transmittance also decreased to approximately 27%. After this, it took another few coating runs with iteratively corrected layer thicknesses to arrive at the currently best result presented in the next section.

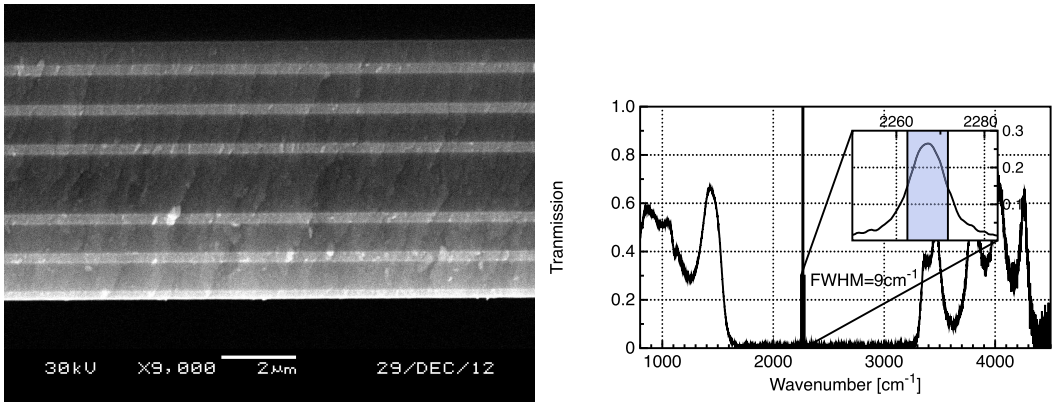


Figure 5.18 *SEM and source-normalized transmission spectrum of the first structure (first-order cavity, three double-layers) grown after the debug process. Inset: Zoom-in on the central transmission peak.*

5.3 Final Design of the Interference Filter

A Fabry-Perot bandpass interference filter is built from alternating quarter-wave layers of high- and low-index transparent materials to result in high transmission only in a narrow band. For the MIR, we use germanium as the high-index material and yttrium fluoride as the low-index material. The optical properties of YF_3 and Ge are plotted in Fig. 5.7 and Fig. 5.6, as measured using the SENDIRA Spectroscopic MIR Ellipsometer at Sentech Instruments on bare 500 nm single-layer samples. The absorption in YF_3 from approximately 2000-3700 cm^{-1} stems from water absorbed by the layer due to its slightly porous microstructure, behavior generally found in evaporated oxides and fluorides, but not in Ge. To avoid this absorption in filters, it is essential to design the filter in such a way that a Ge layer is the outermost layer of the structure, protecting the YF_3 from water contamination.

The filter has a basic single-cavity Fabry-Perot design with a low-index first-order spacer with three layer pairs per reflector. The admittance into the substrate — which being silicon has a high refractive index of approximately 3.4 and subsequently high Fresnel reflectance when bare — was increased by adding a single-layer anti-reflection coating of yttrium oxide with refractive index ≈ 1.7.

The qualitative structure is:

$$B/S/I/HGLGHGLGHGLGLGHGLGHGLGH$$

Where:
I is an initial layer of Y_2O_3 that acts as an anti-reflective (AR) coating
S is the Si substrate (low-doped)
H is a $\lambda/4$ layer of Ge
L is a $\lambda/4$ layer of YF_3
G is a thin adhesion layer of Y_2O_3
B is a backside coating of either Y_2O_3 as AR coating, or a highly reflective coating, to produce a coupled-cavity effect as described in the Section 5.4.

The filter is designed in such a way that it allows tuning in both directions around a central wavelength, thus the design angle is an oblique one, preferably 20° or 30°, but not normal. In this case, the $\lambda/4$ layer thicknesses are determined by the equation,

$$d_i = \frac{\lambda_{\text{design}}}{4\Re(n_i)\cos\left\{\arcsin\left[\sin\left(\theta_{\text{design}}\right)/\Re(n_i)\right]\right\}}$$

where d_i denotes the thickness of each layer type (H, L, or I, for high-index, low-index, and initial layers), n_i is the corresponding complex index of refraction, and θ_{design} is the incident angle for which the filter is designed, i.e. where the transmission wavelength equals λ_{design}. This equation simply reflects the fact that at the design angle, the layer thicknesses are modified using Snell's law to have an optical path length of $\lambda_{\text{design}}/4$.

The transmission wavelength tunes with angle according to [145]

$$\lambda(\theta) = \lambda(0)\sqrt{1 - \sin^2(\theta)/n_{\text{eff}}^2} \tag{5.20}$$

where

$$n_{\text{eff}} \approx n_L/\sqrt{1 - n_L/n_H + (n_L/n_H)^2} \tag{5.21}$$

is the effective index of refraction of the Filter [151].

The filter demonstrated here has a design wavelength $\lambda_{\text{design}} = 4.0\,\mu$m and a design angle $\theta_{\text{design}} = 45°$ and consequently design layer thicknesses of $d_I = 653.4$ nm, $d_H = 252.5$ nm, and $d_L = 763.8$ nm.

5.4 Performance of the Filter

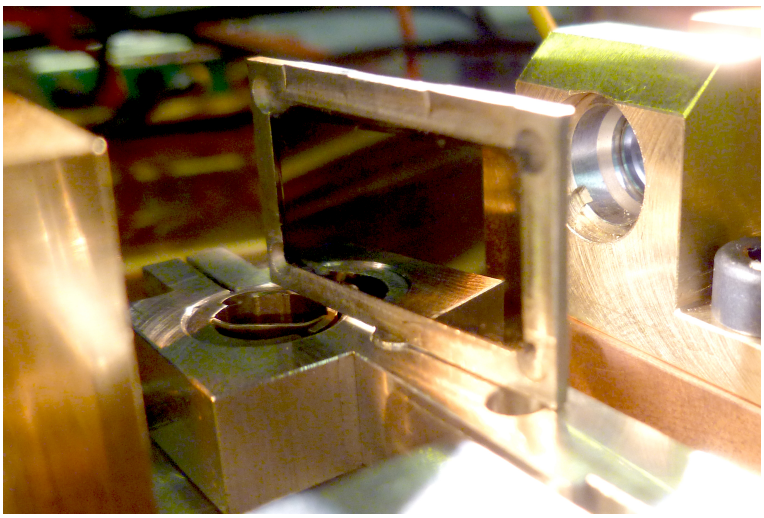


Figure 5.19 *Photograph of the final filter used to tune an External Cavity QCL.*

The filter presented in this section with the design described in the last section has lead to publication [193]. The resulting filter, pictured in Fig. 5.19 appears mirror-like with no visible cracks. There is no change in appearance and performance after 18 months exposure to ambient humidity. It fulfills the requirements for adhesion and abrasion as stated in MIL-C-48497. A cross-sectional SEM micrograph can be seen in Fig. 5.20. As is clear from the image, the layers are very homogenous and the interfaces are smooth and abrupt. The filters were deposited onto substrates that are 10 mm by 20 mm in size.

Fig. 5.21 shows the transmission spectrum as measured with an FTIR (top) and the transmission spectrum as calculated using the design parameters (bottom) at normal incidence which are both in very good agreement with each other. The central peak has a FWHM of 6 cm^{-1}, as can be seen from the inset, which equals approximately 0.3% of the central wavelength of approximately 2280 cm^{-1}. The peak hight in this measurement is approximately 0.4. Both the width and the hight of the central peak in this measurement are not correct. This is due to two reasons. One is that the FTIR's probe beam is not collimated, thus the

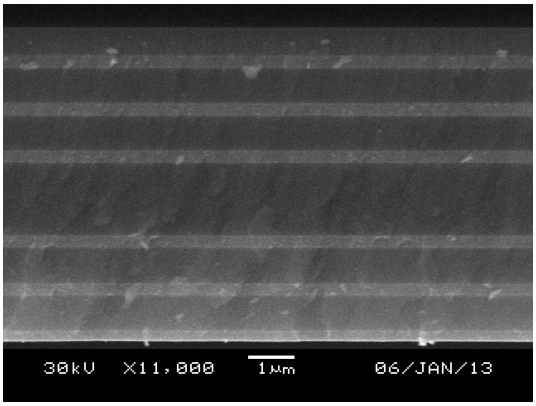


Figure 5.20 *Cross-sectional SEM micrograph of filter structure (substrate towards the top of the image).*

peak is "smeared out" as is the result of the many incident angles in conjunction with the fact that the peak position tunes with angle. This makes the peak both smaller and wider. The second is due to the fact that the backside of the filter is not AR coated in conjunction with the fact that the FTIR's globar source is not coherent. Thus, the peak is expected to be lowered by the amount that is lost to the Fresnel reflection at the backside, which for the high refractive index of silicon is expected to be approximately 28%. All of these effects are avoided in the following by performing subsequent measurements using our S-ECQCL Spectrometer presented in Chapter 4 as a tunable coherent MIR light source with a collimated probe beam. This leads to other effects that are also discussed shortly.

Fig. 5.22 shows a zoomed-in plot of the central transmission peak of the filter obtained using our Littrow ECQCL Spectrometer instead of an FTIR (black dots). Also plotted in Fig. 5.22 are the calculated transmission of the design (dashed red line) with the targeted layer thicknesses. The solid blue line is the calculated transmission using the actual layer thicknesses. The fitted values of the layer thicknesses are $d_H = 250.0$ nm and $d_L = 761.3$ nm and differ from the targeted values ($d_H = 252.5$ nm, and $d_L = 763.8$ nm) by less than 3 nm. Assuming the transparency of the low-index material is the limiting factor of the peak tranmission, the extinction coefficient was calculated to be approximately $k \approx 4.5 \times 10^{-4}$.

The transmission peak of the filter at six different incident angles can be seen in Fig. 5.23 also obtained using our tunable Littrow ECQCL. At normal incidence the transmission is located at 2276 cm^{-1} or 4.40 μm, has an amplitude of 55% and a FWHM of 3.2 cm^{-1} or 6 nm, which is 0.14% of the central wavelength.

The ripples superimposed on the peaks seen in Figs. 5.22 and 5.23 are the Fabry-Perot modes of the substrate due to the backside of the filter not being AR-coated. These tune at a different rate with varying angle than the main peak. This effect deteriorates the tuning behavior and is simply avoided by coating the backside. It can on the other hand be increased by coating the backside with a highly-reflective coating, producing a filter with a 1-2 orders of magnitude smaller bandwidth with comparable peak transmission. This is because the substrate then acts as a second high-finesse cavity with a very small free spectral range and consequently much smaller peak width. Since the peaks can always be tuned with angle so

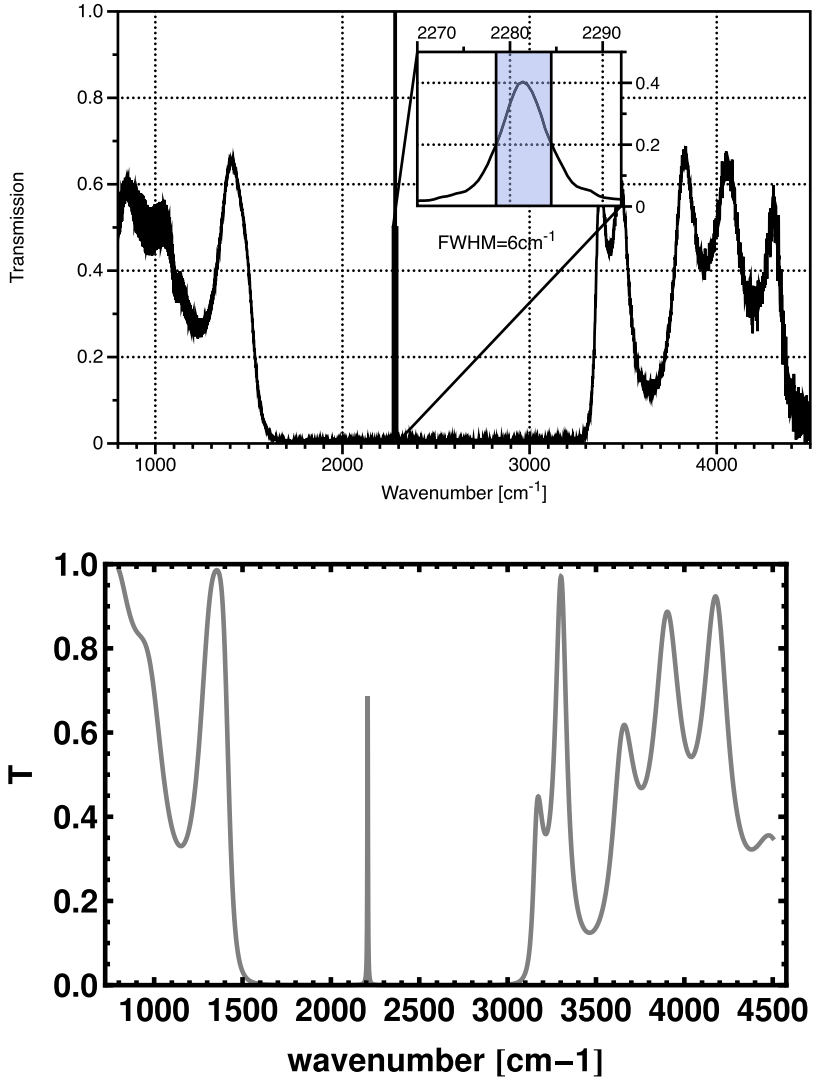


Figure 5.21 *Top: Transmission spectrum as measured with an FTIR. Bottom: Transmission spectrum as calculated using the design parameters.*

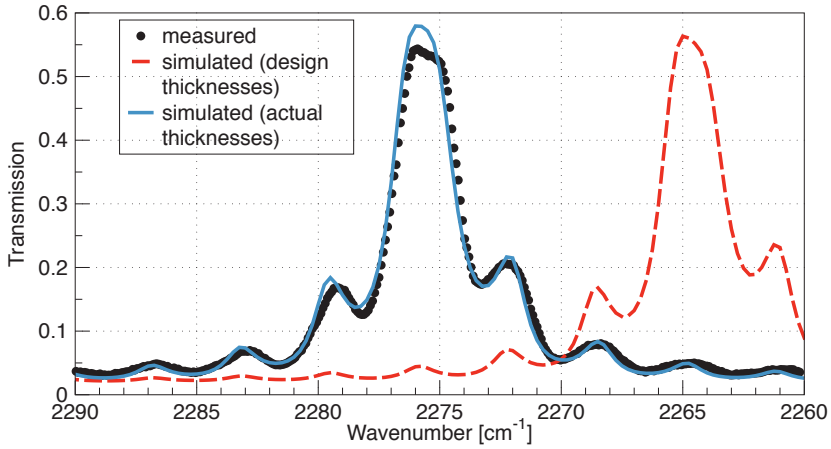


Figure 5.22 *Transmission characteristic of the filter. Dashed red line: Calculated transmission using the discussed design and design layer thicknesses. Black dots: Measured transmission, using our tunable Littrow ECQCL. Solid blue line: Calculated transmission using same design and actual layer thicknesses that differ from the targeted ones by less than 3 nm. The ripples superimposed on the peak are the Fabry-Perot modes of the substrate due to the backside of the filter not being AR-coated. This is explained later in this section.*

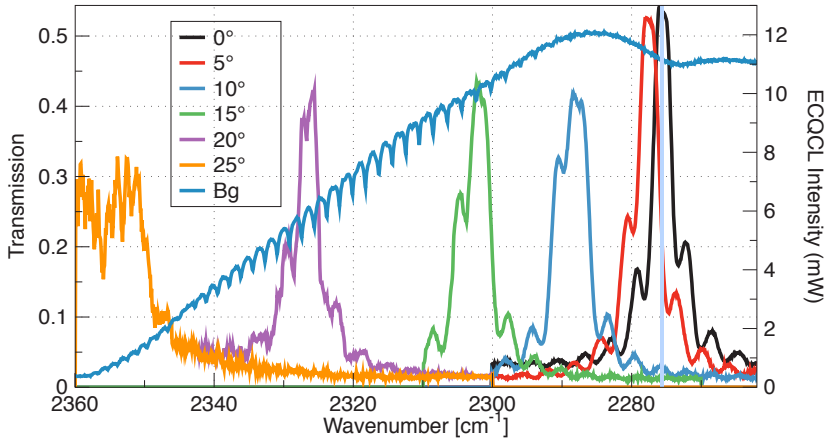


Figure 5.23 *Tuning behavior of the filter, measured using an ECQCL.*

that one coincides with the central wavelength of the main peak (red and black lines in Fig. 5.23), this produces a high-transmission central peak and small satellite peaks in the tails of the transmission range of the main peak, which can be ignored in most cases. This effect can be used to produce miniature ECQCLs with fixed wavelengths or small tuning ranges with highly stable and predictable wavelengths comparable to DFB (Distributed Feedback) QCLs, but without the need to actively stabilize the laser's temperature. Hitting a particular wavelength when developing a new production process is also much easier than with DFB due to the residual tunability.

The coupled-cavity effect only becomes visible when the beam has a coherence length longer than the effective optical path through filter and substrate. The intensity of the probe beam as a function of frequency is also plotted in Fig. 5.23 for reference. Clearly visible are the CO_2 lines of the ambient air and the decreasing intensity towards higher frequencies limiting the useful spectral range at about 2350 cm^{-1}. Consequently, there is considerable noise imposed on the transmission peak at 25°, unfortunately making it impossible to use this particular laser to take spectra at higher angles. But it is to be noted, that the peak amplitude and width do not vary appreciably up until 20°. The smaller maximum transmission at 10°, 15°, and 20° are not due to lower transmission of the Fabry-Perot cavity at more oblique angles, but due to more unfavorable coincidence of the filter and substrate cavities. If we arbitrarily set 50° as a conservative estimate of the upper limit of the incident angle at which the filter works without performance degradation, the tuning range is approximately 300 cm^{-1} or 13% of the central wavelength.

The thickness variations of the deposited layers across one filter is negligible, i.e. in the sub-nanometer range. This can be seen from the following argument. The theoretical and experimental performance in Fig. 5.22 agree well upon illumination of the entire filter area with the probe beam. But an absolute layer thickness shift as small as 2.5 nm shifts the transmission peak by 12 cm^{-1} as is also clear from Fig. 5.22 (difference between red and blue lines). Thus if a variation of similar magnitude was present on a single filter, the effect on the transmission would be catastrophic. Since adjacent filters are only 10 mm apart on the substrate holder in the deposition system, the thickness variation is also marginal across different filters of the same coating run.

The peak position as a function of angle (relative to normal) is plotted in Fig. 5.28. The line is a fit of Eq. (5.20) to the points resulting in an effective index of refraction $n_{\text{eff}} = 1.66$, which is close to the value of 1.65 calculated for the ideal structure using Eq. (5.21).

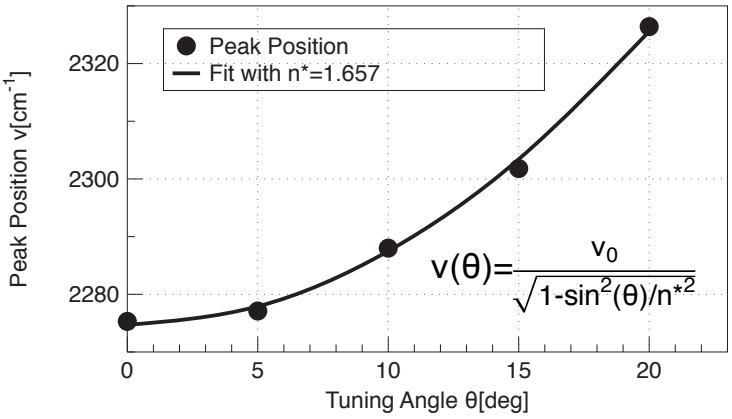


Figure 5.24 *Peak position as a function of angle (relative to normal).*

5.5 Alignment-stabilized filter-tuned External-Cavity QCL

The filter whose design, fabrication, and performance was discussed in the previous sections has been employed in a cat's eye-stabilized, filter-tuned External Cavity QCL. The experimental setup is presented in the following section along with a subsequent discussion of its performance. This setup has lead to publication [194].

5.5.1 Experimental Setup

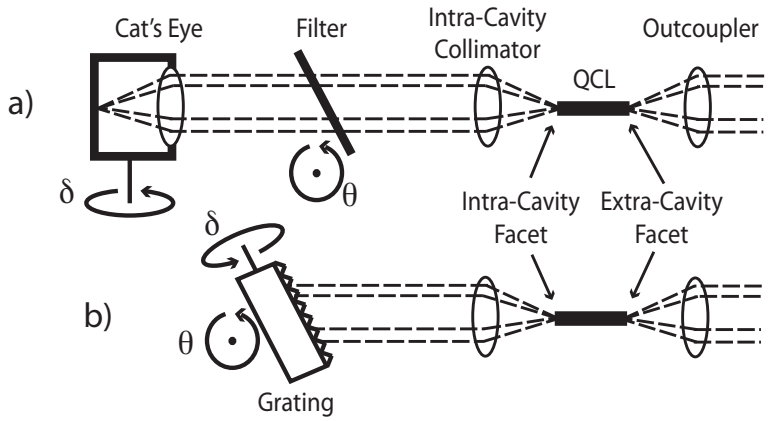


Figure 5.25 *Sketch of the experimental setup of the ECQCL including tuning angles (θ) and detuning angles (δ). a) Setup with cat's eye and filter. b) Reference setup with grating in Littrow configuration.*

The EC setup is sketched in Fig. 5.25 a). For reference, we also assembled a Littrow cavity of the same length, using a 300rules/mm gold-coated aluminum grating with nominal reflectivity of 85%, sketched in Fig. 5.25 b). Figure 5.26 shows a photograph of the filter-tuned setup. The QCL in use is a 32-cascade, 20 μm×6 mm strain-compensated composite-barrier design with 25 mW average output power when driven as an uncoated Fabry-Perot laser with 100-ns pulses at 1% duty cycle and emitting around 4.4 μm. A simple single-layer AR-coating, deposited on one of the facets according to recipes published earlier [192], raises the

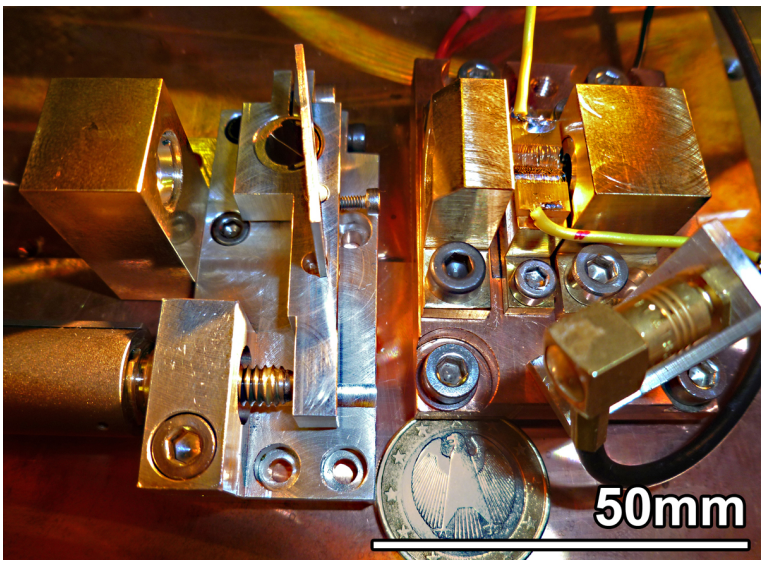


Figure 5.26 *Photograph of the experimental configuration.*

threshold current from 4.50 A to 6.01 A, which corresponds to a residual reflectivity of $\approx 1.5\%$. Collimation of the output and intra-cavity beams was done using AR-coated aspheric lenses with numerical apertures (NA) of 0.85 and 0.56, respectively. A "cat's eye"-type retroreflector consists of a lens and a mirror positioned in its focal plane. The lens used in the cat's eye is identical to the intra-cavity collimator (ICC) and the mirror is a 10 mm-diameter flat gold mirror. Lens and mirror are positioned precisely relative to one another and cemented together for maximum rigidity. The free-space path length of the cavity was ≈23 mm. The collimation was not disassembled between the two setups for better comparison.

To investigate the tuning behavior, both filter and grating were tuned and spectra were taken using a Bruker Equinox 55 FTIR. To investigate the stability of the optical feedback to misalignment, the lasers were misaligned in a controlled manner by consecutively mounting retroreflector and grating on a rotation stage and tilting them by a small angle δ with respect to their perfectly aligned orientation. The decrease of the laser's ability to emit stimulated radiation was then quantified through its increased threshold current as a function of tilt measured with a bolometer. To avoid possible sources of error, the stability measurement of the cat's eye setup was performed without the filter.

5.5.2 Tuning and Spectral Stability

The spectra taken at different filter angles are plotted in Fig. 5.27 as a function of wavenumber $\nu = 1/\lambda$. These were taken at a current of 7 A at room temperature. The FTIR was set to a resolution of 1 cm^{-1}, which is the limiting factor of the plotted linewidth. The actual linewidth is ≈ 0.1 cm^{-1} as is estimated from known chirp values for pulsed QCLs in Littrow cavities [22] but could be orders of magnitude smaller when driven CW.

The positions of the peaks in Fig. 5.27 as a function of tuning angle θ of the filter are plotted in Fig. 5.28 (solid circles), the solid line is a fit of $\nu(\theta) = \nu(0)/(1-\sin^2(\theta)/n_{\mathrm{eff}}^2)^{1/2}$ [145], where

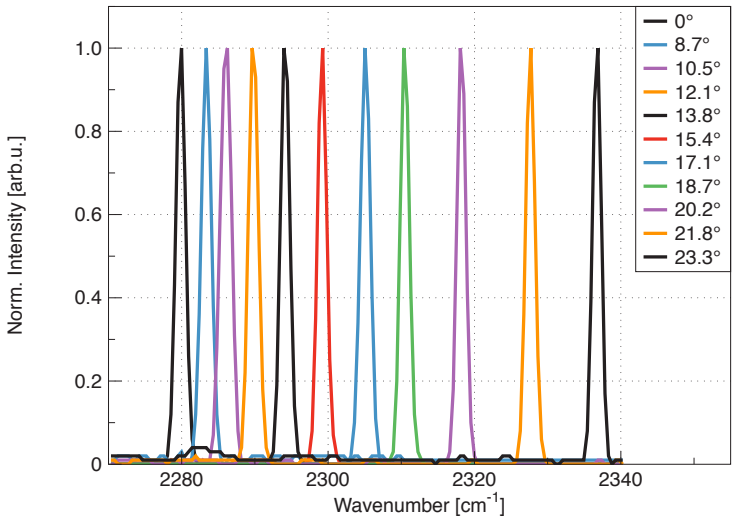


Figure 5.27 *Spectra of filter-tuned ECQCL taken at different angles. The peak width is limited by the resolution of the FTIR.*

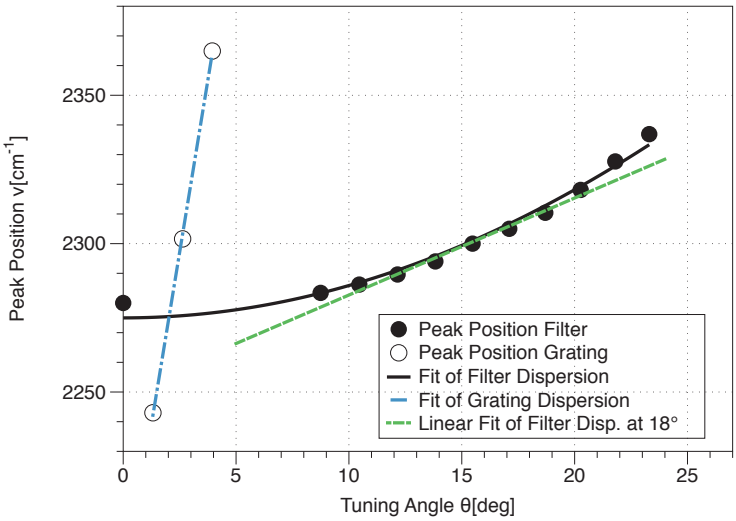


Figure 5.28 *Tuning behavior of filter and grating (for the latter, the angle is to be understood relative to 43.25° and having a negative sign).*

n_{eff} is the effective index of refraction with a fitted value of 1.78. Also plotted in Fig. 5.28 is the tuning behavior of the grating (open circles) for comparison along with a fit line (dashed-dotted) in the form of the *Grating Equation*: $\nu(\theta) = (2d\sin(\theta))^{-1}$, where d is the groove distance of the grating. The incline of the curves at a given angle is the inverse of the angular dispersion and is a measure for the sensitivity of how thermal and mechanical vibration translate into spectral fluctuation. The incline of the grating is $D_{\text{grating}} = 46.4\,\text{cm}^{-1}/\text{deg}$ and the incline of the filter at an exemplary point, at 18°, is $D_{\text{filter}} = 3.27\,\text{cm}^{-1}/\text{deg}$ (dashed line). Thus the spectral stability against vibration of the filter is increased by a factor of $D_{\text{grating}}/D_{\text{filter}} = 14$.

At identical conditions, the power of the filter-tuned setup is reduced to approximately half compared with the grating-tuned setup due to the filter's currently low 55% (single-pass) transmissivity T compared with the grating's nominal reflectivity of 85%, so while the filter setup has an overall feedback of ≈ 19%, the Littrow setup has ≈ 30%. However, a test without the filter shows that the feedback from the cat's eye is approximately identical to the ICC's collection efficiency of ≈ 65%, which, due to the reduction of coupling losses, is over twice as large as the achievable grating feedback, proving ample potential to be realized with better filters. The limiting factor of the tuning range here is the imperfect overlap of the filter's tuning range with the gain of the QCL (2180 to 2380 cm^{-1}) as well as the lower transmissivity. With slight improvement to the filters to reach $T \approx 75\%$, we expect feedback from the filters to be on par with gratings, at which point, given the correct central wavelength, the filter's entire tuning range of ≈10–20% [193] of the central wavelength could be available for laser tuning.

Due to the geometry of the setup, the beam walk-off introduced by the angled filter would have zero effect, if the beam diameter was smaller than the lens apertures by a margin on the order of the filter thickness. Since, in our case, the beam fully illuminates the collimating lens's aperture, the beam walk-off does cause a very slight decrease in collection efficiency, feedback, and output intensity towards large angles. It does not however shift the position or angle of the output beam. In the following, we discuss the increase in threshold current as a function of cat's eye and grating tilt.

5.5.3 Stability of Optical Feedback

The threshold current as a function of the feedback $\bar{F}$ from the grating or cat's eye, respectively, is [16, 147]

$$I_{\text{th}} = \frac{bl(\alpha_w - \ln(R_0\bar{F})/(2l))}{\Gamma g} \tag{5.22}$$

where $R_0 = 0.28$ is the reflectivity of the extra-cavity (out-coupling) QCL facet, $b = 20\,\mu\text{m}$ and $l = 0.6$ cm are the width and length of the stripe, $\alpha_w = 4.1\,\text{cm}^{-1}$ is the distributed waveguide loss, and $\Gamma g = 2.3\,\text{cm/kA}$ is the gain multiplied with the optical confinement factor of the active region. When there is no external reflector, $\bar{F}$ is identical to the reflectivity of the intra-cavity facet. If there is an external reflector, $\bar{F}$ is the overlap of the Gaussian waveguide mode $E_i(z)$ and the reflected Gaussian wave $E_r(z)$ at the position $z = z_0$ of the intra-cavity

facet, $\bar{F} = |\int\int E_i(z_0)^* E_r(z_0) \mathrm{d}x\mathrm{d}y|^2$. The computation of this for the case of a grating at a small misalignment tilt δ is equivalent to the computation of the coupling efficiency of two single-mode fibers [158, 159] at a tilt 2δ and yields

$$\bar{F} = R_\mathrm{g} \exp\left[-\delta^2 \Big/ \left(\frac{\lambda}{\pi\omega_g}\right)^2\right] \tag{5.23}$$

where δ is the angle between the beam impinging on the grating and the plane normal to the grooves. R_g is the nominal (intensity) reflectivity of the grating at the given wavelength λ multiplied by other inevitable coupling losses, and ω_g is the width or spot size of the collimated beam at the position of the grating. Substituting Eq. (5.23) into Eq. (5.22) yields a parabolic dependence of threshold current on grating misalignment with the curvature

$$\frac{\mathrm{d}^2 I_\mathrm{th}}{\mathrm{d}\delta^2} = \frac{b}{\Gamma g}\left(\frac{\pi\omega_g}{\lambda}\right)^2. \tag{5.24}$$

The quantity $(\frac{\mathrm{d}^2 I_\mathrm{th}}{\mathrm{d}\delta^2})^{-1/2}$ or ω_g^{-1} is a measure for misalignment stability [147].

The cat's eye's misalignment sensitivity has been treated before by Zorabedian [147], Baillard [145], and others in the form of the relative movement of the mirror with respect to the lens and other elements. In this case Eq. (5.24) holds for both the grating and the cat's eye, with ω_g being either the spot size on the grating or, alternatively, the beam waist, ω_0, of the focused beam on the mirror of the cat's eye. Thus the stability is inversely proportional to the spot size on the mirror or grating. While the spot size on the grating can not be arbitrarily small, because the bandwidth of the grating is inversely proportional to the number of illuminated grooves, the spot size within the cat's eye is not subject to such constrictions, and is only determined by the lens and the wavelength (in Gaussian optics) through

$$\omega_0 \approx f\lambda/(\pi\omega_{ic}), \tag{5.25}$$

where f is the focal distance of the lens and ω_{ic} is the intra-cavity beam width. In case the lens is perfectly illuminated by the beam, Eq. (5.25) becomes $\omega_0 \approx \lambda/(\pi\mathrm{NA})$, where NA is the numerical aperture of the focussing lens.

Since a cat's eye can be fabricated as a compact module by positioning the mirror precisely in the focal plane of the small lens and rigidly cementing the two together, we have chosen to treat the more realistic case where the cat's eye as a whole is tilted. This is equivalent to treating the reflection of an inclined beam from an untilted cat's eye. With this treatment, we find the stability to misalignment is still greatly increased compared with a grating, but not as much as in the idealized case of Eq. (5.24) as we discuss in the following.

In the paraxial ray and the thin lens approximations, when the focal distance of the lens is a constant, f, the cat's eye is a perfect retroreflector and the beam returns on its incident path with no losses and the only divergence is due to the finite spot size of the Gaussian beam. However, at larger tilt angles, these approximations break down and the focal distance becomes a function of incline [160] and also splits up for tangential and sagittal rays due to

astigmatism. The expression for the tangential ray becomes

$$f(\delta) = -\sqrt{1/n^2 - \sin(\delta)^2}\,\frac{R\sqrt{n^2 - \sin(\delta)^2}}{\cos(\delta) - \sqrt{n^2 - \sin(\delta)^2}}, \tag{5.26}$$

where n is the refractive index of the (plano-convex) lens and R its radius of curvature. The lens in use has $R = 6.52$ mm and $n = 2.63$ so that at a tilt angle of 10°, the focal length reduces from 4mm to 3.52 mm. When using a flat mirror, the beam waist walks out of the mirror plane for larger tilt angles. This results in an increasing beam divergence angle with increasing tilt angle after traversing the lens upon return. This in turn leads to an increase in coupling losses back into the QCL's facet.

This decrease of focal distance can be countered by using a mirror with just the right curvature, so that the beam waist of the focussed beam remains on the mirror's surface in which case the stability reaches the waist-limited value, determined by Eqs. (5.24) and (5.25). Then, estimating a waist of $\omega_0 \approx 3.8\,\mu$m, assuming $\omega_{ic} = \omega_g \approx 1.5$ mm, a focal length of 4 mm, and wavelength of $4.4\,\mu$m, the setup would be $\omega_g/\omega_0 \approx 395$ times more stable to misalignment than for a grating.

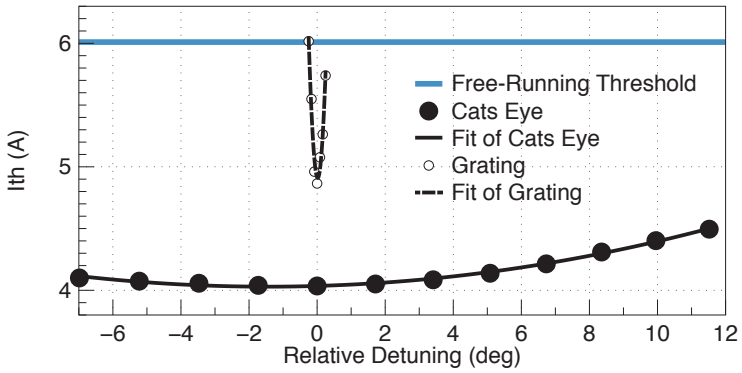


Figure 5.29 *Misalignment sensitivity of filter and grating. Lasing threshold current as a function of angular detuning of cat's eye and of Littrow grating.*

The measured stability in the form of the threshold current as a function of angular detuning of the cat's eye and the Littrow grating are plotted in Figs. 5.29 and 5.30 along with the free-running threshold, which is due to the residual reflectivity of the AR-coated facet.

The parabolas in Fig. 5.29 and Fig. 5.30 were fitted in the form $I_{\mathrm{th}} = I^0 + 1/2 I''\delta^2$ with $I^0_{\mathrm{grating}} = 4.91$ A, $I^0_{\mathrm{cateye}} = 4.03$ A, $I''_{\mathrm{grating}} = 32.05$ A/deg^2 and $I''_{\mathrm{cateye}} = 5.60 \times 10^{-3}$ A/deg^2. The reduction in misalignment sensitivity $(I''_{\mathrm{grating}}/I''_{\mathrm{cateye}})^{1/2}$ is 76, which is nearly two orders of magnitude but not as large as 395 if the correct mirror curvature was used.

With the use of Eq. (5.24), I''_{grating} yields a spot size ω_g =1.5 mm on the grating which is in perfect agreement with the previous assumption. However, the "effective" spot size on the cat's eye's mirror, ω_0, as calculated from Eq. (5.24), is 20 μm and not 3.8 μm, but this value is overestimated due to the focal distance reduction.

While the maximum feedback F of the cat's eye (without filter) is $\approx$ 65%, as can be calculated from Eq. (5.22) using $I^0_{\mathrm{cateye}} = 4.03$ A, the feedback max for the grating in this

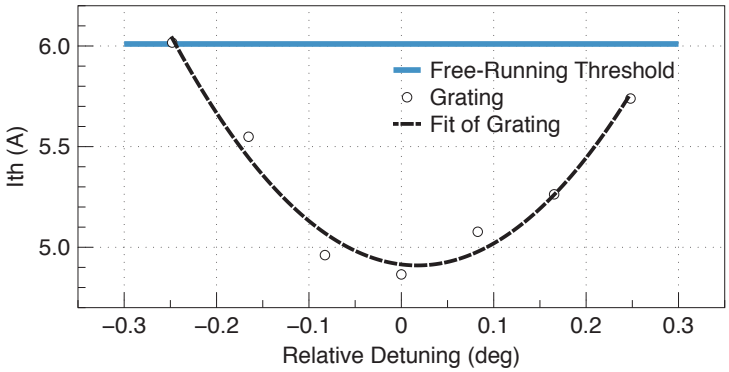


Figure 5.30 *Misalignment sensitivity of grating. Lasing threshold current as a function of angular detuning of Littrow grating.*

measurement is only 17%.

6

Conclusion and Summary

This thesis thoroughly investigates theoretically and experimentally the effects many physical parameters have on the performance of EC-QCLs. These include, among others, the anti-reflection coating, the type of optics, and the geometrical as well as mechanical and structural properties of the EC setup. This was done by assembling three very different EC configurations and comparing and discussing their performance, as well as advantages and disadvantages for different purposes using mainly QCLs from the same original wafer for better comparability.

In Chapter 2 we discussed that the anti-reflection coating of the QCL facet is one of the key elements of any External-Cavity that substantially determines the behavior of the entire laser. The ideal anti-reflection coating perfectly eliminates the reflectivity of the QCL facet. Realistic anti-reflection coatings have a residual reflectivity on the order of a tenth of a percent to several percent. This leads to coupled-cavity effects that lead to mode-hopping and intensity fluctuations with the period of the Fabry-Perot modes of the chip as the laser is tuned.

We have discussed a theoretical model to calculate the transmission and reflection characteristic of a multi-layer AR coating stack, have solved it analytically for the one and two-layer case and introduced a Matrix Method for easy numerical computation of arbitrary layer structures. This model can be used to design optimal AR coatings using a given material structure.

We have pointed out that even outside an External Cavity configuration, the anti-reflection coating substantially alters the behavior of a QCL, as it – in the ideal case – opens the laser resonator on one side. Due to spontaneous emission of the active region, depending on pump rate the laser now behaves like a light-emitting diode, a super-luminescent diode, or a diode laser. This behavior makes it difficult to interpret the P-I curves to calculate a residual reflectivity and with it the quality of the coating. We have pointed out that the literature normally uses one of two ways to determine the residual reflectivity of a facet coating, either from the ratio of the threshold currents before and after coating, or from the ratio of the slope efficiencies of the light emitted from the coated and uncoated sides of the laser after coating. These approaches ignore the effect of spontaneous emission, and for very low facet reflectivities, result in vastly different values and are consequently not

trust-worthy. We have developed and discussed a theoretical model that treats the QCL waveguide as an amplifier for traveling photons, similar to fiber amplifiers and lasers, and use a phenomenological approach to model the spontaneous emission. Through self-consistency and consistency with experiment we are able to fit the residual reflectivity of the AR-coated facet to a value, which we believe is the best estimate for the true value.

Finally, we have developed and discussed a setup to quantify the residual reflectivity in-situ during rf-magnetron sputter deposition of the coating.

In Chapter 3 we have demonstrated the setup, alignment, and operation of a very versatile research-oriented Littrow-type External Cavity QCL (V-ECQCL) that can be used for many different types of semiconductor laser regardless of wavelength range and power level.

The setup is done using a visible pilot beam on a specialized optical bench on which the detachable External Cavity is mounted. The procedure unlinks the mechanical degrees of freedom and therefore these can be adjusted and fixed one by one. Through this, the procedure is oblivious to the light emitted from the semiconductor laser and can be used even for experimental lasers with low power or the need for cooling inside a cryostat.

We have characterized and analyzed many aspects of the operation including the geometrical parameters of the setup and how they relate to the actual output of the laser light in terms of power and modal structure. This was done using a theoretical coupled-cavity model using the experimental values as input parameters. In-depth analysis was done on the example of an 5-mm QCL emitting at 11 μm and coated with an AR coating with residual reflectivity of approximately 9% inside an EC with 9-cm free-space path length. The 150-rules/mm grating in conjunction with a 1" beam resulted in a grating bandwidth that spans approximately 3 EC modes and half of the FP mode distance. This allowed us to resolve not only the periodical near-vanishing of the intensity with the period of the FP mode spacing, but also a super-oscillation that results from the interaction of EC length and FP modes. On this example we have demonstrated how to track a single mode with appropriate variation of the EC length as the grating is tuned and the chip temperature is varied periodically. We have also shown experimentally and theoretically how the effect diminishes with better AR coatings on the example of a 6-mm QCL emitting at 4.6 μm with an AR coating that has a reflectivity of 1.5%. The latter suppressed the mode hop-related intensity fluctuations to the limit of detector noise, the former had remaining fluctuations of approximately 75% when there is no mode-tracking employed.

Mode-hop related intensity fluctuations can also be eliminated in pulsed operation by utilizing sufficiently long QCLs to have the mode spacing comparable to the wavelength chirp, with the tradeoff of a larger linewidth as discussed on the example of a 10-mm QCL. Digital tuning on the External Cavity modes can be made acceptable by working with long External Cavity path lengths.

We have demonstrated the use of this EC-QCL for gas absorption in conjunction with a photoacoustic cell. The versatility of this spectrometer was demonstrated on absorption measurements on ambient and exhaled carbon dioxide in the wavelength range around 4.35 μm and ammonia in the wavelength range around 11.1 μm. The tuning ranges are 48 cm^{-1} and

67 cm^{-1}, respectively.

When using photoacoustic detection, the sensitivity can be maximized by decreasing the diameter of the probing beam, since the PA signal is proportional to the intensity and not the power of probing light. Alternatively, placing an optical power meter next to the output facet and using the external cavity itself as a multi-pass absorption cell delivers good results.

In Chapter 4 we have described and experimentally investigated the design strategy, setup, performance, and first applications of a mechanically stabilized External Cavity QCL spectrometer (S-ECQCL). The design strategy achieves ruggedness through decoupling the adjustment and fixation of each mechanical degree of freedom, using only custom-made parts for structure in a very compact and robust design.

The finished modular spectrometer consist of a hermetically sealed laser head with a thermoelectrically cooled QCL (flushed with an inert gas), a controller housing, a hand-held controller, and a computer. The laser head contains a water-cooled heat sink designed from first principles with over 400 W/K thermal conductance on an area of 5x5 cm^2 using 3 parallel coolant channels with a length of 28 cm and a cross-section of 8 mm^2, optimized for a volumetric flow of 4 L/min. The spectrometer is fully software-controlled using a program based on a Queued State Machine programming concept. The software is used to gather, process, and display the spectrum in conjunction with an external detector. There is great freedom of experimental design as well as the flexibility to integrate the spectrometer into larger experiments through multiple trigger inputs and outputs, gating, and syncing.

Do to its mechanical stability there is a high level of experimental reproducibility, making the setup ideal to investigate the emission under different operating conditions. The tuning range, tuning envelopes, and P-I curves have been investigated for three different duty cycles of the driving current in pulsed operation at two different heat sink temperatures using the same 6-mm QCL emitting at around 4.4 μm with an AR coating with a residual reflectivity of 1.5% used in previous chapters. The pulse duration was 100 ns at repetition rates of 10, 87, and 200 kHz, resulting in duty cycles of 0.1, 0.87, and 2%. The temperatures were 21°C and -16°C.

The characteristic temperature T_0 of the uncoated QCL, as determined from the threshold of the P-I curves at the two temperatures was approximately 100 K, with no clear dependence on duty cycle. The characteristic temperature T_0 for the coated QCL in the EC configuration with the grating tuned to the gain maximum, was approximately 90 K, also with no clear dependence on duty cycle.

The P-I curve of the cooled, AR-coated QCL without feedback from the grating shows behavior of a laser with a clearly defined threshold and rollover, but at room temperature, the QCL behaves like a super-luminescent diode with a P-I characteristic curved upwards with no clear threshold. Also, the cooled, AR-coated QCL without feedback from the grating exhibits a characteristic of a saturated upper laser state resulting from strong filling paired with weak relaxation channels resulting in a very distinct rollover characteristic of the P-I curve.

The maximum pulse output power of the External Cavity – i.e. at the maximum pump rate

that does not mix FP oscillation into the emission across the entire tuning range – was nearly independent of both temperature and duty cycle, in good agreement with basic theoretical considerations. Its value was approximately 850 mW. The power spectra do not resemble a single clipped Lorentzan, as would be expected from a single intersubband transition, but rather a spectrum of 4-5 transitions. The (root-mean-square) intensity fluctuations due to mode hopping were approximately 4% of the intensity for both cooled and uncooled EC-QCLs. The tuning ranges varied slightly do to slightly different relative pumping at the different temperatures, but were all in the ranges of (180-210) cm^{-1} or approximately 9% of the central wavelength.

This EC-QCL-based spectrometer has been used for preliminary high-resolution absorption measurements on ambient water vapor with a QCL emitting around 5.7 μm and on isotopologue-resolved measurements of CO_2. It has also been used to measure with high resolution the transmission characteristics of the filters discussed in Chapter 5. Finally, this spectrometer has been shipped to a collaborating research group to perform pump-probe measurements on dyes.

In Chapter 5 we have described the design, fabrication and experimental performance of a bandpass interference filter for the MIR that is suitable for tuning a self-aligned external-cavity QCL (F-ECQCL). Using the materials and techniques described, filters with narrow transmission bands and good mechanical stability can be produced for the 3–12 μm range.

The filter design is an unblocked first-order Fabry-Perot cavity using yttrium fluoride as the low-index material, germanium as the high-index material and silicon as the substrate that is anti-reflection coated with a layer of yttrium oxide. The layers were deposited using electron beam evaporation with highly optimized deposition parameters and surface treatment.

A demonstration filter, with transmission band at about 2300 cm^{-1} and a band width of 3.2 cm^{-1} has peak transmission of 55% and can be angle-tuned through at least 300 cm^{-1}. The extinction coefficient of the low-index material was calculated to be approximately $k \approx 4.5 \times 10^{-4}$. The filter fulfills the requirements for adhesion and abrasion as stated in MIL-C-48497.

This filter has been used to demonstrate and experimentally and theoretically investigate a filter-tuned EC-QCL, stabilized with a "cat's eye"-type retroreflector made of a lens and a mirror positioned at it's focal plane. Compared with a grating-tuned setup, the stability to misalignment of this setup is increased by a factor of 76 and the spectral stability is increased by a factor of 14 at an exemplary point (18° tuning angle). The stability to misalignment can be increased to approximately 400 times that of the grating setup by using a mirror with the correct curvature.

While with the best possible collimation using a 3-mm diameter lens with a numerical aperture of 0.85 the peak feedback from the grating is only approximately $(17-30)\%$, however with a near 100% transmission filter, the feedback can be increased to approximately 65%.

Aside from the increased mechanical and spectral stability, the demonstrated setup has the advantages that the optical components are mass-producible, the mechanical components are relieved of their very strict accuracy and precision requirements, and the setup is miniaturiz-

able, since the beam diameter can be shrunk almost arbitrarily.

Finally, in the Appendix we give a thorough theoretical review of light-matter interactions in order to provide a framework upon which it can be decided from first principles how to choose the right infrared materials to use for high-Q band-pass interference filters as discussed in Chapter 5.

Bibliography

1. MASSELINK, W., Semtsiv, M. & Kischkat, J. *Lasereinrichtung* DE Patent App. DE201, 210,000,038. July 2013.
2. Kosterev, A. *et al.* Application of quantum cascade lasers to trace gas analysis. *Applied Physics B: Lasers and Optics* **90,** 165–176 (2008).
3. Wang, C. & Sahay, P. Breath analysis using laser spectroscopic techniques: Breath biomarkers, spectral fingerprints, and detection limits. *Sensors* **9,** 8230–8262 (2009).
4. Griffiths, P. R. & Haseth, J. A. D. *Fourier Transform Infrared Spectrometry* 2nd ed. (Wiley-Blackwell, Hoboken, New Jersey, 2007).
5. Tittel, F., Richter, D. & Fried, A. Mid-infrared laser applications in spectroscopy. *Solid-State Mid-Infrared Laser Sources,* 458–529 (2003).
6. Hugi, A. *et al.* External cavity quantum cascade laser tunable from 7.6 to 11.4 mu m. *Applied Physics Letters* **95.** doi:10.1063/1.3193539 (2009).
7. Yao, Y. *et al.* Broadband quantum cascade laser gain medium based on a "continuum-to-bound" active region design. *Applied Physics Letters* **96.** doi:10.1063/1.3431577 (2010).
8. Faist, J. *et al.* Distributed feedback quantum cascade lasers. *Applied physics letters* **70,** 2670–2672 (1997).
9. Kazarinov, R. F. & Suris, R. A. Possibility of Amplification of Electromagnetic Waves in a Semiconductor with a Superlattice. *Soviet Physics Semiconductors - USSR* **5,** 797–800 (1971).
10. Faist, J. *et al.* Quantum cascade laser. *Science* **264,** 553 (1994).
11. Semtsiv, M. *et al.* Above room temperature operation of short wavelength (λ= 3.8 μm) strain-compensated In 0.73 Ga 0.27 As–AlAs quantum-cascade lasers. *Applied physics letters* **85,** 1478–1480 (2004).
12. Semtsiv, M., Wienold, M., Dressler, S. & Masselink, W. Short-wavelength (λ 3.05 μm) InP-based strain-compensated quantum-cascade laser. *Applied physics letters* **90,** 051111–051111 (2007).
13. Faist, J. *Quantum Cascade Lasers* (Oxford Univ Press, Oxford, UK, 2013).
14. Sirtori, C. in (ed Paiella, R.) chap. Quantum Cascade Lasers: Overview of Basic Principles of Operation and State of the Art (McGraw-Hill, New York, 2006).
15. Cockburn, J. English. in *Mid-infrared Semiconductor Optoelectronics* (ed Krier, A.) 323–355 (Springer London, 2006). doi:10.1007/1-84628-209-8_10.
16. Gmachl, C., Capasso, F., Sivco, D. & Cho, A. Recent progress in quantum cascade lasers and applications. *Reports on progress in physics* **64,** 1533 (2001).
17. Faist, K. *et al.* Bound-to-continuum and two-phonon resonance, quantum-cascade lasers for high duty cycle, high-temperature operation. *Quantum Electronics, IEEE Journal of* **38,** 533–546 (2002).

18. Yao, Y., Hoffman, A. J. & Gmachl, C. F. Mid-infrared quantum cascade lasers. *Nature Photonics* **6,** 432–439 (2012).

19. Le, H. *et al.* Broad wavelength tunability of grating-coupled external cavity midinfrared semiconductor lasers. *Applied physics letters* **69,** 2804–2806 (1996).

20. Le, H. *et al. External cavity mid-infrared semiconductor lasers* in *Proc. SPIE* **3001** (SPIE, Bellingham, 1997), 298–308.

21. Luo, G. P. *et al.* Grating-tuned external-cavity quantum-cascade semiconductor lasers. *Applied Physics Letters* **78,** 2834–2836 (2001).

22. Luo, G. P. *et al.* Broadly wavelength-tunable external cavity mid-infrared quantum cascade lasers. *Ieee Journal of Quantum Electronics* **38,** 486–494 (2002).

23. Totschnig, G., Winter, F., Pustogov, V., Faist, J. & Muller, A. Mid-infrared external-cavity quantum-cascade laser. *Optics Letters* **27,** 1788–1790 (2002).

24. Peng, C., Luo, G. P. & Le, H. Q. Broadband, continuous, and fine-tune properties of external-cavity thermoelectric-stabilized mid-infrared quantum-cascade lasers. *Applied Optics* **42,** 4877–4882 (2003).

25. Peng, C., Zhang, H. L. & Le, H. Q. Mid-infrared external-cavity two-segment quantum-cascade laser. *Applied Physics Letters* **83,** 4098–4100 (2003).

26. Maulini, R., Beck, M., Faist, J. & Gini, E. Broadband tuning of external cavity bound-to-continuum quantum-cascade lasers. *Applied Physics Letters* **84,** 1659–1661 (2004).

27. Maulini, R., Yarekha, D. A., Bulliard, J. M., Giovannini, M. & Faist, J. Continuous-wave operation of a broadly tunable thermoelectrically cooled external cavity quantum-cascade laser. *Optics Letters* **30,** 2584–2586 (2005).

28. Wysocki, G. *et al.* Widely tunable mode-hop free external cavity quantum cascade laser for high resolution spectroscopic applications. *Applied Physics B-Lasers and Optics* **81,** 769–777 (2005).

29. Wysocki, G. *et al.* Widely tunable mode-hop free external cavity quantum cascade lasers for high resolution spectroscopy and chemical sensing. *Applied Physics B-Lasers and Optics* **92,** 305–311 (2008).

30. Pushkarsky, M., Tsekoun, A., Dunayevskiy, I. G., Go, R. & Patel, C. K. N. Sub-parts-per-billion level detection of NO2 using room-temperature quantum cascade lasers. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 10846–10849 (2006).

31. Patel, C. K. N. Laser photoacoustic spectroscopy helps fight terrorism: High sensitivity detection of chemical Warfare Agent and explosives. *European Physical Journal-Special Topics* **153,** 1–18 (2008).

32. Weida, M. J., Arnone, D. & Day, T. Tunable QC laser up mid-IR opens sensing applications. *Laser Focus World* **42,** 109–110 (2006).

33. Takeuchi, E., Thomas, K. & Day, T. Applications multiply for external-cavity QCLs. *Laser Focus World* **45,** 83–86 (2009).

34. Weida, M. J. & Yee, B. Quantum cascade laser based replacement for FTIR microscopy. *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues Ix* **7902.** doi:10.1117/12.873954 (2011).

35. Caffey, D. *et al.* Recent Results from Broadly Tunable External Cavity Quantum Cascade Lasers. *Novel in-Plane Semiconductor Lasers X* **7953.** doi:10.1117/12.875093 (2011).

36. Fotheringham, E., Weida, M. J., Chapman, W. B. & Day, T. Progress towards compact broadly tunable laser modules for high-resolution mid-IR spectroscopy and commercial applications. *Quantum Sensing and Nanophotonic Devices Ix* **8268.** doi:10.1117/12.914414 (2012).

37. Gensty, T., von Staden, J. & Elsasser, W. Investigations of the intensity noise of quantum cascade lasers. *Tm-Technisches Messen* **72,** 380–387 (2005).

38. Mohan, A. *et al.* Room-temperature continuous-wave operation of an external-cavity quantum cascade laser. *Optics Letters* **32,** 2792–2794 (2007).

39. Mukherjee, N., Go, R. & Patel, K. N. Linewidth measurement of external grating cavity quantum cascade laser using saturation spectroscopy. *Applied Physics Letters* **92.** doi:10.1063/1.2901038 (2008).

40. Maulini, R. *et al.* Widely tunable high-power external cavity quantum cascade laser operating in continuous-wave at room temperature. *Electronics Letters* **45,** 107–U30 (2009).

41. Hinkov, B. *et al.* Time-resolved characterization of external-cavity quantum-cascade lasers. *Applied Physics Letters* **94.** doi:10.1063/1.3142875 (2009).

42. Knabe, K. *et al.* Frequency characterization of a swept- and fixed-wavelength external-cavity quantum cascade laser by use of a frequency comb. *Optics Express* **20,** 12432–12442 (2012).

43. Leonhaeuser-Rein, B., Drzewietzki, L., Schad, F. & Elsaesser, W. Investigations of mode beat frequencies of external cavity semiconductor lasers. *Applied Physics B-Lasers and Optics* **113,** 215–220 (2013).

44. Inoue, T., Tsushima, K., Mori, S. & Kasahara, K. Quantum cascade laser intensity noise under external feedback conditions estimated from self-mixing method. *Electronics Letters* **49,** 407–408 (2013).

45. Tan, S. *et al.* Low-threshold, high SMSR tunable external cavity quantum cascade laser around 4.7 mu m. *Optical and Quantum Electronics* **45,** 1147–1155 (2013).

46. Weidmann, D. & Wysocki, G. High-resolution broadband (> 100 cm(-1)) infrared heterodyne spectro-radiometry using an external cavity quantum cascade laser. *Optics Express* **17,** 248–259 (2009).

47. Weidmann, D., Tsai, T., Macleod, N. A. & Wysocki, G. Atmospheric observations of multiple molecular species using ultra-high-resolution external cavity quantum cascade laser heterodyne radiometry. *Optics Letters* **36,** 1951–1953 (2011).

48. Fuchs, F. *et al.* Imaging standoff detection of explosives using widely tunable mid-infrared quantum cascade lasers. *Optical Engineering* **49.** doi:10.1117/1.3506195 (2010).

49. Fuchs, F. *et al.* Imaging stand-off detection of explosives using tunable MIR quantum cascade lasers. *Quantum Sensing and Nanophotonic Devices Vii* **7608.** doi:10.1117/12.840464 (2010).

50. Hinkov, B. *et al.* Time-resolved spectral characteristics of external-cavity quantum cascade lasers and their application to stand-off detection of explosives. *Applied Physics B-Lasers and Optics* **100,** 253–260 (2010).

51. Phillips, M. C., Myers, T. L., Wojcik, M. D. & Cannon, B. D. External cavity quantum cascade laser for quartz tuning fork photoacoustic spectroscopy of broad absorption features. *Optics Letters* **32,** 1177–1179 (2007).

52. Phillips, M. C. *et al.* Design and performance of a sensor system for detection of multiple chemicals using an external cavity quantum cascade laser. *Quantum Sensing and Nanophotonic Devices Vii* **7608.** doi:10.1117/12.842120 (2010).

53. Phillips, M. C. & Taubman, M. S. Intracavity sensing via compliance voltage in an external cavity quantum cascade laser. *Optics Letters* **37,** 2664–2666 (2012).

54. Phillips, M. C. & Ho, N. Infrared hyperspectral imaging using a broadly tunable external cavity quantum cascade laser and microbolometer focal plane array. *Optics Express* **16,** 1836–1845 (2008).

55. Stupar, D. *et al.* Fully reflective external-cavity setup for quantum-cascade lasers as a local oscillator in mid-infrared wavelength heterodyne spectroscopy. *Applied Optics* **47,** 2993–2997 (2008).

56. Sonnabend, G., Wirtz, D. & Schieder, R. Evaluation of quantum-cascade lasers as local oscillators for infrared heterodyne spectroscopy. *Applied Optics* **44,** 7170–7172 (2005).

57. Lewicki, R., Wysocki, G., Kosterev, A. A. & Tittel, F. K. QEPAS based detection of broadband absorbing molecules using a widely tunable, cw quantum cascade laser at 8.4 mu m. *Optics Express* **15,** 7357–7366 (2007).

58. Kachanov, A., Koulikov, S. & Tittel, F. K. Cavity-enhanced optical feedback-assisted photo-acoustic spectroscopy with a 10.4 mu m external cavity quantum cascade laser. *Applied Physics B-Lasers and Optics* **110,** 47–56 (2013).

59. Lewicki, R., Doty, I. James H., Curl, R. F., Tittel, F. K. & Wysocki, G. Ultrasensitive detection of nitric oxide at 5.33 mu m by using external cavity quantum cascade laser-based Faraday rotation spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 12587–12592 (2009).

60. Dong, L., Spagnolo, V., Lewicki, R. & Tittel, F. K. Ppb-level detection of nitric oxide using an external cavity quantum cascade laser based QEPAS sensor. *Optics Express* **19,** 24037–24045 (2011).

61. Gong, L. *et al.* Atmospheric ammonia measurements in Houston, TX using an external-cavity quantum cascade laser-based sensor. *Atmospheric Chemistry and Physics* **11,** 9721–9733 (2011).

62. Dong, L. *et al.* Ultra-sensitive carbon monoxide detection by using EC-QCL based quartz-enhanced photoacoustic spectroscopy. *Applied Physics B-Lasers and Optics* **107,** 275–283 (2012).

63. Spagnolo, V., Kosterev, A. A., Dong, L., Lewicki, R. & Tittel, F. K. NO trace gas sensor based on quartz-enhanced photoacoustic spectroscopy and external cavity quantum cascade laser. *Applied Physics B-Lasers and Optics* **100,** 125–130 (2010).

64. Patimisco, P. *et al.* Coupling external cavity mid-IR quantum cascade lasers with low loss hollow metallic/dielectric waveguides. *Applied Physics B-Lasers and Optics* **108,** 255–260 (2012).

65. Spagnolo, V. *et al.* Mid-infrared fiber-coupled QCL-QEPAS sensor. *Applied Physics B-Lasers and Optics* **112,** 25–33 (2013).

66. Patimisco, P. *et al.* Low-Loss Hollow Waveguide Fibers for Mid-Infrared Quantum Cascade Laser Sensing Applications. *Sensors* **13,** 1329–1340 (2013).

67. Brandstetter, M., Genner, A., Anic, K. & Lendl, B. Tunable external cavity quantum cascade laser for the simultaneous determination of glucose and lactate in aqueous phase. *Analyst* **135,** 3260–3265 (2010).

68. Brandstetter, M. & Lendl, B. Tunable mid-infrared lasers in physical chemosensors towards the detection of physiologically relevant parameters in biofluids. *Sensors and Actuators B-Chemical* **170,** 189–195 (2012).

69. Brandstetter, M., Volgger, L., Genner, A., Jungbauer, C. & Lendl, B. Direct determination of glucose, lactate and triglycerides in blood serum by a tunable quantum cascade laser-based mid-IR sensor. *Applied Physics B-Lasers and Optics* **110,** 233–239 (2013).

70. Brandstetter, M. *et al.* Reagent-free monitoring of multiple clinically relevant parameters in human blood plasma using a mid-infrared quantum cascade laser based sensor system. *Analyst* **138,** 4022–4028 (2013).

71. Young, C. *et al.* External cavity widely tunable quantum cascade laser based hollow waveguide gas sensors for multianalyte detection. *Sensors and Actuators B-Chemical* **140,** 24–28 (2009).

72. Hancock, G., van Helden, J. H., Peverall, R., Ritchie, G. A. D. & Walker, R. J. Direct and wavelength modulation spectroscopy using a cw external cavity quantum cascade laser. *Applied Physics Letters* **94.** doi:10.1063/1.3141521 (2009).

73. Van Helden, J. H., Peverall, R., Ritchie, G. A. D. & Walker, R. J. Rapid passage effects in nitrous oxide induced by a chirped external cavity quantum cascade laser. *Applied Physics Letters* **94.** doi:10.1063/1.3079420 (2009).

74. Walker, R. J., van Helden, J. H. & Ritchie, G. A. D. Quantum cascade laser absorption spectroscopy of the 1 <- 0 band of deuterium bromide at 5 mu m. *Chemical Physics Letters* **501,** 20–24 (2010).

75. Karpf, A. & Rao, G. N. Absorption and wavelength modulation spectroscopy of NO2 using a tunable, external cavity continuous wave quantum cascade laser. *Applied Optics* **48,** 408–413 (2009).

76. Karpf, A. & Rao, G. N. Enhanced sensitivity for the detection of trace gases using multiple line integrated absorption spectroscopy. *Applied Optics* **48,** 5061–5066 (2009).

77. Karpf, A. & Rao, G. N. Enhancement of trace gas detection by integrating wavelength modulated spectra across multiple lines. *Applied Optics* **49,** 1406–1413 (2010).

78. Rao, G. N. & Karpf, A. Extremely sensitive detection of NO2 employing off-axis integrated cavity output spectroscopy coupled with multiple-line integrated absorption spectroscopy. *Applied Optics* **50,** 1915–1924 (2011).

79. Rao, G. N. & Karpf, A. High sensitivity detection of NO2 employing cavity ringdown spectroscopy and an external cavity continuously tunable quantum cascade laser. *Applied Optics* **49,** 4906–4914 (2010).

80. Chao, X., Jeffries, J. B. & Hanson, R. K. In situ absorption sensor for NO in combustion gases with a 5.2 mu m quantum-cascade laser. *Proceedings of the Combustion Institute* **33,** 725–733 (2011).

81. Chao, X., Jeffries, J. B. & Hanson, R. K. Wavelength-modulation-spectroscopy for real-time, in situ NO detection in combustion gases with a 5.2 mu m quantum-cascade laser. *Applied Physics B-Lasers and Optics* **106,** 987–997 (2012).

82. Kottmann, J., Rey, J. M., Luginbuehl, J., Reichmann, E. & Sigrist, M. W. Glucose sensing in human epidermis using mid-infrared photoacoustic detection. *Biomedical Optics Express* **3,** 667–680 (2012).

83. Kottmann, J., Grob, U., Rey, J. M. & Sigrist, M. W. Mid-Infrared Fiber-Coupled Photoacoustic Sensor for Biomedical Applications. *Sensors* **13,** 535–549 (2013).

84. Pleitez, M. A. *et al.* In Vivo Noninvasive Monitoring of Glucose Concentration in Human Epidermis by Mid-Infrared Pulsed Photoacoustic Spectroscopy. *Analytical Chemistry* **85,** 1013–1020 (2013).

85. Wen, Q. & Michaelian, K. H. Mid-infrared photoacoustic spectroscopy of solids using an external-cavity quantum-cascade laser. *Optics Letters* **33,** 1875–1877 (2008).

86. Sydoryk, I., Lim, A., Jaeger, W., Tulip, J. & Parsons, M. T. Detection of benzene and toluene gases using a midinfrared continuous-wave external cavity quantum cascade laser at atmospheric pressure. *Applied Optics* **49,** 945–949 (2010).

87. Parsons, M. T. *et al.* Real-time monitoring of benzene, toluene, and p-xylene in a photoreaction chamber with a tunable mid-infrared laser and ultraviolet differential optical absorption spectroscopy. *Applied Optics* **50,** A90–A99 (2011).

88. Woerle, K. *et al.* Breath Analysis with Broadly Tunable Quantum Cascade Lasers. *Analytical Chemistry* **85,** 2697–2702 (2013).

89. Suter, J. D., Bernacki, B. & Phillips, M. C. Spectral and angular dependence of mid-infrared diffuse scattering from explosives residues for standoff detection using external cavity quantum cascade lasers. *Applied Physics B-Lasers and Optics* **108,** 965–974 (2012).

90. Knabe, K. *et al.* Absolute spectroscopy of N2O near 4.5 mu m with a comb-calibrated, frequency-swept quantum cascade laser spectrometer. *Optics Express* **21,** 1020–1029 (2013).

91. Kole, M. R., Reddy, R. K., Schulmerich, M. V., Gelber, M. K. & Bhargava, R. Discrete Frequency Infrared Microspectroscopy and Imaging with a Tunable Quantum Cascade Laser. *Analytical Chemistry* **84,** 10366–10372 (2012).

92. Luedeke, S., Pfeifer, M. & Fischer, P. Quantum-Cascade Laser-Based Vibrational Circular Dichroism. *Journal of the American Chemical Society* **133,** 5704–5707 (2011).

93. Ribaudo, T., Adams, D. C., Passmore, B., Shaner, E. A. & Wasserman, D. Spectral and spatial investigation of midinfrared surface waves on a plasmonic grating. *Applied Physics Letters* **94.** doi:10.1063/1.3140569 (2009).

94. Pushkarsky, M. B. *et al.* High-sensitivity detection of TNT. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 19630–19634 (2006).

95. Mukherjee, A. *et al.* Sub-parts-per-billion level detection of dimethyl methyl phosphonate (DMMP) by quantum cascade laser photoacoustic spectroscopy. *Applied Optics* **47,** 1543–1548 (2008).

96. Mukherjee, N. & Patel, C. K. N. Molecular fine structure and transition dipole moment of NO(2) using an external cavity quantum cascade laser. *Chemical Physics Letters* **462,** 10–13 (2008).

97. Maulini, R., Mohan, A., Giovannini, M., Faist, J. & Gini, E. External cavity quantum-cascade laser tunable from 8.2 to 10.4 mu m using a gain element with a heterogeneous cascade. *Applied Physics Letters* **88.** doi:10.1063/1.2205183 (2006).

98. Wittmann, A., Hugi, A., Gini, E., Hoyler, N. & Faist, J. Heterogeneous High-Performance Quantum-Cascade Laser Sources for Broad-Band Tuning. *Ieee Journal of Quantum Electronics* **44,** 1083–1088 (2008).

99. Bismuto, A., Riedi, S., Hinkov, B., Beck, M. & Faist, J. Sb-free quantum cascade lasers in the 3-4 mu m spectral range. *Semiconductor Science and Technology* **27.** doi:10.1088/0268-1242/27/4/045013 (2012).

100. Riedi, S., Hugi, A., Bismuto, A., Beck, M. & Faist, J. Broadband external cavity tuning in the 3-4 mu m window. *Applied Physics Letters* **103.** doi:10.1063/1.4813851 (2013).

101. Dougakiuchi, T. *et al.* Broadband Tuning of External Cavity Dual-Upper-State Quantum-Cascade Lasers in Continuous Wave Operation. *Applied Physics Express* **4.** doi:10.1143/apex.4.102101 (2011).

102. Mroziewicz, B. External cavity wavelength tunable semiconductor lasers-a review. *Opto-Electronics Review* **16,** 347–366 (2008).

103. Zhang, H. L. *et al.* External-cavity tunable mid-infrared laser using off-band surface-emitting Bragg grating coupler. *Applied Physics Letters* **86.** doi:10.1063/1.1885188 (2005).

104. Tsai, T. & Wysocki, G. External-cavity quantum cascade lasers with fast wavelength scanning. *Applied Physics B-Lasers and Optics* **100,** 243–251 (2010).

105. Tsai, T. & Wysocki, G. Active wavelength control of an external cavity quantum cascade laser. *Applied Physics B-Lasers and Optics* **109,** 415–421 (2012).

106. De Labachelerie, M. & Passedat, G. Mode-hop suppression of Littrow grating-tuned lasers. *Applied optics* **32,** 269–274 (1993).

107. De Labachelerie, M., Sasada, H. & Passedat, G. Mode-hop suppression of Littrow grating-tuned lasers: erratum. *Applied optics* **33,** 3817–3819 (1994).

108. Tormen, M., Peter, Y.-A., Niedermann, P., Hoogerwerf, A. & Stanley, R. Deformable MEMS grating for wide tunability and high operating speed. *Journal of Optics a-Pure and Applied Optics* **8,** S337–S340 (2006).

109. Malara, P. *et al.* External ring-cavity quantum cascade lasers. *Applied Physics Letters* **102.** doi:10.1063/1.4800073 (2013).

110. Wasserman, D., Shaner, E. A. & Cederberg, J. G. Midinfrared doping-tunable extraordinary transmission from sub-wavelength gratings. *Applied Physics Letters* **90.** doi:10.1063/1.2737138 (2007).

111. Lee, B. G. *et al.* DFB quantum cascade laser arrays. *Quantum Electronics, IEEE Journal of* **45,** 554–565 (2009).

112. Lee, B. G. *et al.* Broadband distributed-feedback quantum cascade laser array operating from 8.0 to 9.8 m. *IEEE Photonics Technology Letters* **21** (2009).

113. Lee, B. G. *et al.* Widely tunable single-mode quantum cascade laser source for mid-infrared spectroscopy. *Applied Physics Letters* **91,** 231101 (2007).

114. Capasso, F. High-performance midinfrared quantum cascade lasers. *Optical Engineering* **49.** doi:10.1117/1.3505844 (2010).

115. Lee, B. G. *et al.* Beam combining of quantum cascade laser arrays. *Optics Express* **17,** 16216–16224 (2009).

116. Hugger, S. *et al.* Power scaling of quantum cascade lasers via multiemitter beam combining. *Optical Engineering* **49.** doi:10.1117/1.3498766 (2010).

117. Bloom, G. *et al.* Passive coherent beam combining of quantum-cascade lasers with a Dammann grating. *Optics Letters* **36,** 3810–3812 (2011).

118. Maulini, R. *Broadly tunable mid-infrared quantum cascade lasers for spectroscopic applications* PhD thesis (Institut de Physique Université de Neuchatel, 2006).

119. Hugi, A., Maulini, R. & Faist, J. External cavity quantum cascade laser. *Semiconductor Science and Technology* **25.** doi:10.1088/0268-1242/25/8/083001 (2010).

120. Tittel, F. K. *et al.* in *Quantum Sensing and Nanophotonic Devices V* (eds Sudharsanan, R. & Jelen, C.) Z9000–Z9000 (2008). doi:10.1117/12.754722.

121. Sigrist, M. W. Trace gas monitoring by laser photoacoustic spectroscopy and related techniques (plenary). *Review of Scientific Instruments* **74,** 486–490 (2003).

122. Risby, T. H. & Tittel, F. K. Current status of midinfrared quantum and interband cascade lasers for clinical breath analysis. *Optical Engineering* **49.** doi:10.1117/1.3498768 (2010).

123. Rao, G. N. & Karpf, A. External cavity tunable quantum cascade lasers and their applications to trace gas monitoring. *Applied Optics* **50,** A100–A115 (2011).

124. Brandstetter, M., Genner, A., Anic, K. & Lendl, B. Tunable Mid-IR Lasers: A New Avenue To Robust And Versatile Physical Chemosensors. *Eurosensors Xxiv Conference* **5,** 1001–1004 (2010).

125. Roepcke, J., Davies, P. B., Lang, N., Rousseau, A. & Welzel, S. Applications of quantum cascade lasers in plasma diagnostics: a review. *Journal of Physics D-Applied Physics* **45.** doi:10.1088/0022-3727/45/42/423001 (2012).

126. Wojcik, M. D. *et al.* in *Optically Based Biological and Chemical Sensing for Defence* (eds Carrano, J. C. & Zukauskas, A.) 154–165 (2004). doi:10.1117/12.601503.

127. Lotem, H. Mode-hop suppression of Littrow grating-tuned lasers: comment. *Applied optics* **33,** 3816–3816 (1994).

128. Weinstein, W. The reflectivity and transmissivity of multiple thin coatings. *JOSA* **37,** 576–581 (1947).

129. Abeles, F. Recherches sur la propagation des ondes electromagnetiques sinusoidales dans les milieus stratifies. Applications aux couches minces. *Annales de physique* **5,** 596–640 (1950).

130. Abeles, F. Recherches sur la propagation des ondes electromagnetiques sinusoidales dans les milieus stratifies. Applications aux couches minces. *Annales de physique* **5,** 706–782 (1950).

131. Heavens, O. Optical properties of thin films. *Reports on Progress in Physics* **23,** 1 (1960).

132. Musset, A. & Thelen, A. Multilayer Antireflection Coatings. *Progress in Optics* **8,** 201–237 (1970).

133. Saleh, B., Teich, M. & Saleh, B. *Fundamentals of photonics* (Wiley Online Library, Hoboken, New Jersey, 1991).

134. Xu, J. *et al.* Tunable terahertz quantum cascade lasers with an external cavity. *Applied Physics Letters* **91,** 121104–121104 (2007).

135. Gorelik, A. *et al.* Miniaturized resonant photoacoustic cell of inclined geometry for trace-gas detection. *Applied Physics B* **100,** 283–289 (2010).

136. Fedosenko, O. *et al.* Scaling the output power of quantum-cascade lasers with a number of cascades. *Journal of Crystal Growth* **323,** 484–487 (2011).

137. *HITRAN on the Web* http://hitran.iao.ru. Downloaded: 14.12.2014.

138. *G2 Engineering Technical Note* http://www.g2-engineering.com/technology-kinematics. Downloaded: 20.11.2014.

139. *Riverhawk Technical Note* http://flexpivots.com/double-ended-pivot-bearing/. Downloaded: 10.11.2014.

140. De Nevers, N. *Fluid Mechanics* (Addison-Wesley, Boston, MA, 1970).

141. Moody, L. F. Friction factors for pipe flow. *Trans. Asme* **66,** 671–684 (1944).

142. Idelchik, I. E. Handbook of hydraulic resistance. *Washington, DC, Hemisphere Publishing Corp., 1986, 662 p. Translation.* **1** (1986).

143. Rowe, M. Measurements and computations of flow in pipe bends. *Journal of Fluid Mechanics* **43,** 771–783 (1970).

144. Company, B. W̃., Stultz, S. C. & Kitto, J. B. *Steam: its generation and use* (Babcock & Wilcox, 1992).

145. Baillard, X. *et al.* Interference-filter-stabilized external-cavity diode lasers. *Optics Communications* **266,** 609–613 (2006).

146. Littman, M. & Metcalf, H. Spectrally narrow pulsed dye laser without beam expander. *Applied Optics* **17,** 2224–2227 (1978).

147. Zorabedian, P. & Trutna, W. R. Interference-Filter-Tuned, Alignment-Stabilized, Semiconductor External-Cavity Laser. *Optics Letters* **13,** 826–828 (Oct. 1988).

148. Macleod, A. The early days of optical coatings. *Journal of Optics A: Pure and Applied Optics* **1,** 779 (1999).

149. Hawkins, G., Sherwood, R. & Djotni, K. *Mid-infrared filters for astronomical and remote sensing instrumentation* in *Optical Systems Design* (2008), 710114–710114.

150. Cockburn, J. in (ed Krier, A.) 323–355 (Springer, 2006).

151. Macleod, H. A. *Thin-Film Optical Filters, Fourth Edition* (CRC Press, Taylor and Francis Group, Boca Raton, 2010).

152. Black, P. & Wales, J. Materials for use in the fabrication of infrared interference filters. *Infrared Physics* **8,** 209–222 (1968).

153. Kruschwitz, J. D. T. & Pawlewicz, W. T. Optical and durability properties of infrared transmitting thin films. *Appl. Opt.* **36,** 2157–2159 (Apr. 1997).

154. Pellicori, S. & Colton, E. Fluoride compounds for IR coatings. *Thin Solid Films* **209,** 109–115 (1992).

155. *Fabry-Perot Etalon* Left: Author Kevin J Morse, Right: Author Bob Mellish. Licensed under the Creative Commons Attribution-ShareAlike 3.0 License.

156. *Tydex Optics Technical Note* http://www.tydexoptics.com/pdf/Germanium.pdf. Downloaded: 14.04.2015.

157. Liu, M.-C., Lee, C.-C., Kaneko, M., Nakahira, K. & Takano, Y. Microstructure related properties of lanthanum fluoride films deposited by molybdenum boat evaporation at 193 nm. *Thin Solid Films* **492,** 45–51 (2005).

158. Joyce, W. & DeLoach, B. Alignment of Gaussian beams. *Applied optics* **23,** 4187–4196 (1984).

159. Marcuse, D. Loss analysis of single-mode fiber splices. *Bell Syst. Tech. J* **56,** 703–718 (1977).

160. Massey, G. & Siegman, A. Reflection and refraction of Gaussian light beams at tilted ellipsoidal surfaces. *Applied Optics* **8,** 975–978 (1969).

161. Dresselhaus, M. *Solid State Phsics, Part II, Optical Properties of Solids* http://web.mit.edu/course/6/6.732/www/6.732-pt2.pdf. Downloaded: 15.02.2014.

162. Bekefi, G. *Electromagnetic vibrations, waves, and radiation* (Mit Press, 1977).

163. Jackson, J. D. *Classical Electrodynamics* 3rd ed. (Wiley-VCH, Weinheim, 1998).

164. Lifshitz, E., Pitaevskii, L. & Landau, L. *Electrodynamics of Continuous Media* (Butterworth-Heinemann, Oxford, UK, 1984).

165. Griffiths, D. J. *Introduction to Electrodynamics* (Prentice Hall, Upper Saddle River, NJ, 1999).

166. Born, M. & Wolf, E. *Principals of Optics* (Pergamon Press, New York, 1980).

167. Hecht, E. *Optics* 4th ed. (Addison Wesley, Boston, MA, 1998).

168. Nolting, W. *Grundkurs Theoretische Physik 3: Elektrodynamik* 7th ed. (Springer, Berlin Heidelberg, 2004).

169. Pedrotti, F. L. & Pedrotti, L. S. *Introduction to Optics* (Prentice-Hall, Englewood Cliffs, NJ, 1993).

170. Panofsky, W. K. & Phillips, M. *Classical Electricity and Magnetism* 2nd ed. (Courier Dover Publications, Mineola, NY, 2012).

171. Zangwill, A. *Modern Electrodynamics* (Cambridge University Press, Cambridge, UK, 2012).

172. Greiner, W. & Soff, S. *Classical Electrodynamics (Classical Theoretical Physics)* (Springer, Berlin, 1998).

173. Fließbach, T. *Lehrbuch zur Theoretischen Physik 2, Elektrodynamik* (Springer, Berlin, 2009).

174. Meyer-Arendt, J. R. *Introduction to Classical and Modern Optics* 3rd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1989).

175. Sellmeier, W. III. Ueber die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die erstern, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* **221,** 399–421 (1872).

176. Sellmeier, W. II. Ueber die durch Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* **221,** 520–549 (1872).

177. Sellmeier, W. Ueber die durch die Aetherschwingungen erregten Mitschwingungen der Körpertheilchen und deren Rückwirkung auf die ersteren, besonders zur Erklärung der Dispersion und ihrer Anomalien. *Annalen der Physik* **223,** 386–403 (1872).

178. Sellmeier, W. Ueber die durch die Aetherschwingungen erregten Mitschwingungen der Koerpertheilchen und deren Rueckwirkung auf die ersteren, besonders zur Erklaerung der Dispersion und ihrer Anomalien; *Annalen der Physik* **223,** 525–554 (1872).

179. Helmholtz, H. Zur Theorie der anomalen Dispersion. *Annalen der Physik* **230,** 582–596 (1875).

180. Drude, P. Optische Eigenschaften und Elektronentheorie. *Annalen der Physik* **319,** 677–725 (1904).

181. Drude, P. Optische Eigenschaften und Elektronentheorie. *Annalen der Physik* **319,** 936–961 (1904).

182. Feynman, R. P., Leighton, R. B. & Sands, M. *The Feynman Lectures on Physics* (Addison-Wesley Longman, Amsterdam, 2013).

183. Jenkins, F. A. & White, H. E. *Fundamentals of Optics* 3rd ed. (McGraw-Hill, New York, 1957).

184. Sargent III, M., Scully, M. & Lamb Jr, W. *Laser Physics* 56 (Addison-Wesley, Reading, MA, 1974).

185. Demtröder, W. *Experimentalphysik 2: Elektrizität und Optik* 5th ed. (Springer, Berlin, 2009).

186. Gerthsen, C. *Physik* 24th ed. (ed Meschede, D.) (Springer, Berlin, 2010).

187. Shiles, E., Sasaki, T., Inokuti, M. & Smith, D. Self-consistency and sum-rule tests in the Kramers-Kronig analysis of optical data: applications to aluminum. *Physical Review B* **22,** 1612 (1980).

188. Palik, E. & Ghosh, G. *Handbook of optical constants of solids* (Academic press, New York, 1998).

189. Fujiwara, H. *Spectroscopic ellipsometry: principles and applications* (Wiley, New York, 2007).

190. Tompkins, H. & Irene, E. *Handbook of ellipsometry* (William Andrew, Norwich, New York, 2005).

191. Tompkins, H. & McGahan, W. *Spectroscopic ellipsometry and reflectometry: a user's guide* (Wiley New York, New York, 1999).

Own Publications

192. Kischkat, J. *et al.* Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride. *Applied Optics* **51,** 6789–6798 (2012).

193. Kischkat, J. *et al.* Ultra-narrow angle-tunable Fabry–Perot bandpass interference filter for use as tuning element in infrared lasers. *Infrared Physics & Technology* **67,** 432–435 (2014).

194. Kischkat, J. *et al.* Alignment-Stabilized Interference Filter-Tuned External-Cavity Quantum Cascade Laser. *Optics Letters* **39** (2014).

195. Kischkat, J. *et al. Design, fabrication, and applications of ultra-narrow infrared bandpass interference filters* in *SPIE Security+ Defence* (2013), 889614–889614.

Appendices

A

Dielectric Properties of Solids

A.1 Introduction

Electromagnetic fields influence charged particles through the Lorentz force. This simple underlying concept governs a multitude of processes found in real-world interactions, when "light" of the various frequency regimes – (quasi-) static to X-ray – meets charged particles assembled to the highly complex structure that is the solid. In some cases the particles will start oscillating at the frequency of the light field, thus emitting an electromagnetic field of their own. In other cases, light can be absorbed to change the quantum state of the charged particle. This light in turn can or can not be re-emitted at the same or a different frequency. In general though, the resulting light field – to first order the superposition of the original field and the reemitted fields less the absorbed fields – will differ from the original in intensity, polarization, direction, phase relation, and spectral composition, and even the frequency if non-linear processes are in play. This fact can be utilized to study matter through the use of light, or to purposely alter light through the use of matter.

It is the goal of this chapter to explore the different effects that can contribute to this interaction and thus how the different electronic and structural properties of the elements and their solids translate into macroscopic optical properties, to finally examine the different areas of the periodic system for their suitability as optical materials in the mid-infrared range of the electromagnetic spectrum in search of materials that are "well-suited" to use in the fabrication of mid-infrared optical filters, as well as to define the term "well-suited". Although the dielectric properties of solids is a standard topic covered in many textbooks, this chapter attempts to be a compact reference to be of some use to researchers first choosing materials to produce specialized infrared optical coatings. It also contains what we believe to be a new simplified general theory of the refractivity of transparent solids, see Section A.3.1.1.

The contributions to this interaction can generally be categorized as *dispersive* and *absorptive*, or *parametric* and *non-parametric*, respectively. Parametric processes are elastic scattering processes that leave the internal quantum state of the medium unchanged, whereas absorptive processes don't. Some macroscopic phenomena are caused by parametric processes only, like the diffraction and refraction of light and it will be shown later that they are related

to Rayleigh scattering of individual atoms. Others, like the attenuation of the total light as it traverses the solid are due to non-parametric processes only. While yet others, like reflection at a surface are due to a combination of both. Some processes can be treated very well with classical models, some are inherently quantum-mechanical in nature. Some are properties of the atoms that build the solid, but most are inherently collective. Typically, each interaction has a characteristic energy (or frequency) range that is a result of the binding forces and inertia of the interacting particles or subset of particles, but many different effects can overlay within the same range. It will also be shown how the different contributions, although confined in energy range, intricately influence the response of the system at all energies, and how non-parametric processes leave their trace in the parametric response and vice-versa.

All these effects are fully compiled into the response function of the solid to electromagnetic waves, the dielectric permittivity, which has a real and an imaginary part that will be shown to reflect the two classes of interactions individually. The extraordinary power of this function lies in the fact that it fully masks all the tricky internal granularity of the solid with its 10^{23} particles and numerous convoluted interactions and allows the solid to be treated as a continuum with no other relevant properties than this complex-valued, normally slowly spatially varying scalar function of energy (in case anisotropies are not relevant). It thereby significantly simplifies the overall optical problem to a boundary value problem, where only the interfaces between mostly homogenous media need to be considered.

A.2 The Basic Framework

The following is a basic derivation as given for instance in [161].

The "microscopic" Maxwell equations describe how charges and currents cause electric and magnetic fields, $\vec{E}$ and $\vec{B}$, respectively. In SI-units they read:

$$\nabla \cdot \vec{E} = \frac{\rho_{\text{total}}}{\varepsilon_0}, \tag{A.1}$$

$$\nabla \cdot \vec{B} = 0, \tag{A.2}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{A.3}$$

$$\nabla \times \vec{B} = \vec{j}_{\text{total}} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}. \tag{A.4}$$

Eq. (A.1) states that the sources and drains of electric field lines are charges and Eq. (A.2) states that no "charges" exist for the magnetic field lines ("no magnetic monopoles"), thus these are closed loops. Eq. (A.3) states that electric field lines circle around time-varying magnetic fields and Eq. (A.4) states that magnetic fields circle around electric currents and time-varying electric fields. In this formulation, all charges, ρ_{total}, and currents, $\vec{j}_{\text{total}}$, need to be accounted for, which when considering a solid with $\sim 10^{23}$ atoms per cm^3, each with their

respective nuclei and numerous electrons, is very obviously unfeasible. This is particularly true for the currents when considering all the motions that the electrons and nuclei can undergo, causing a myriad of microscopical currents.

The (very successful) attempt of simplifying this problem for the case of solids stems from the idea that most of the microscopic currents and charges cancel out macroscopically, leaving behind, if anything, only a magnetization density $\vec{M}$ and a polarization density $\vec{P}$. This is executed by splitting up the charges and currents into "free" and "bound" charges and currents and expressing the macroscopically averaged bound charges and currents in therms of the macroscopic quantities $\vec{M}$ and $\vec{P}$.

This ultimately yields an averaged current

$$\bar{j} = j_{\text{free}} + \dot{\vec{P}} + \nabla \times \vec{M} \tag{A.5}$$

and an averaged charge

$$\bar{\rho} = \rho_{\text{free}} + \nabla \cdot \vec{P}. \tag{A.6}$$

With the use of two auxiliary fields, called the displacement field $\vec{D}$ and the magnetizing field $\vec{H}$, defined as

$$\vec{D}(\vec{r}, t) = \varepsilon_0 \vec{E}(\vec{r}, t) + \vec{P}(\vec{r}, t) \tag{A.7}$$

and

$$\vec{H}(\vec{r}, t) = \frac{1}{\mu_0} \vec{B}(\vec{r}, t) - \vec{M}(\vec{r}, t), \tag{A.8}$$

respectively, one finally arrives at the "macroscopic" Maxwell equations, which read (in SI-units):

$$\nabla \cdot \vec{D} = \rho_{\text{free}}, \tag{A.9}$$

$$\nabla \cdot \vec{B} = 0, \tag{A.10}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{A.11}$$

$$\nabla \times \vec{H} = \vec{j}_{\text{free}} + \frac{\partial \vec{D}}{\partial t}. \tag{A.12}$$

These equations look identical in form and shape to the microscopic equations, but instead of explicitly using *all* the charges and currents, these only use the *free* charges and currents, i.e. charges in excess to the neutral solid, and current carried by free carriers only. All the complications of the solid are compounded into the macroscopic auxiliary fields $\vec{D}$ and $\vec{H}$.

Equations (A.9) – (A.9) are still exact, only the book keeping on the currents and charges

has changed. The full complexity of the problem is preserved and lies in the fact that now there are four fields in use, some of which still have arbitrarily complicated structures. They can however be greatly simplified. The ansatz

$$\vec{P} = \varepsilon_0 \chi_e \vec{E} \tag{A.13}$$

and

$$\vec{M} = \chi_m \vec{H}, \tag{A.14}$$

states that the polarization density is proportional to the electric field and the magnetization density to the magnetizing field with proportionality functions χ_e and χ_m. This approximation is valid for "linear" media in the linear regime, where external electric fields (due to, for instance, light impinging on a solid) are very weak compared with the internal electric fields between electrons and nuclei. The terms linear media and linear regime are in fact defined by the validity of Eqs. (A.13) and (A.14), and are surprisingly general, exceptions being, for instance, ferromagnetic materials, so-called non-linear crystals, or very strong fields caused by high-power lasers. In general χ_e and χ_m are tensors, and $\vec{E}$ is not necessarily parallel to $\vec{P}$, and $\vec{H}$ not necessarily to $\vec{M}$. In isotropic media however, they are, and χ_e and χ_m are scalar functions, that normally vary only slowly with position, except at interfaces between materials. They do however still have a very complicated time dependence.

Using Eqs. (A.13) and (A.14) and defining the quantities

$$\varepsilon = \varepsilon_0(1 + \chi_e) = \varepsilon_0 \varepsilon_r \tag{A.15}$$

and

$$\mu = \mu_0(1 + \chi_m) = \mu_0 \mu_r \tag{A.16}$$

and inserting them into Eqs. (A.17) and (A.18), immediately leads to the very simple expressions for $\vec{D}$ and $\vec{H}$

$$\vec{D}(\vec{r}, t) = \varepsilon \vec{E}(\vec{r}, t) \tag{A.17}$$

and

$$\vec{H}(\vec{r}, t) = \frac{1}{\mu} \vec{B}(\vec{r}, t). \tag{A.18}$$

Using an ansatz analog to Eq. (A.13) for the free current,

$$\vec{j}_{\text{free}} = \tilde{\sigma} \vec{E}, \tag{A.19}$$

and standard vector identities, it can be shown that the Macroscopic Maxwell Equations (A.9) – (A.12) have the same solutions as the much simpler wave equations

$$\nabla^2 \vec{E} = \varepsilon_0 \varepsilon_r \mu_0 \mu_r \frac{\partial^2 \vec{E}}{\partial t^2} + \mu_0 \mu_r \sigma \frac{\partial \vec{E}}{\partial t} \tag{A.20}$$

and

$$\nabla^2 \vec{H} = \varepsilon_0 \varepsilon_r \mu_0 \mu_r \frac{\partial^2 \vec{H}}{\partial t^2} + \mu_0 \mu_r \sigma \frac{\partial \vec{H}}{\partial t}, \tag{A.21}$$

which have identical shape for $\vec{E}$ and $\vec{H}$. Since this work deals mostly with non-magnetic materials, as optics often does, we will set

$$\mu_r \approx 1 \tag{A.22}$$

and will be solving Eq. (A.20) for $\vec{E}$, ignoring $\vec{H}$ for simplicity.

Since Eq. (A.20) has the form of a wave equation, it is also clear how to deal with the aforementioned complicated time dependence of the problem: Solutions to Eq. (A.20) can be expressed as a superposition of plane waves

$$\vec{E}(\vec{r}, t) = \vec{E}_0 e^{i(\vec{\kappa} \cdot \vec{r} - \omega t)}, \tag{A.23}$$

with a complex amplitude vector $\vec{E}_0$, a (complex) wave vector $\vec{\kappa}$, and frequency ω. This reduces Eq. (A.20) to the algebraic equation

$$\kappa^2 = \varepsilon_0 \varepsilon_r \mu_0 \mu_r \omega^2 + i \mu_0 \mu_r \sigma \omega = \varepsilon_0 \tilde{\varepsilon}_r \mu_0 \mu_r \omega^2 \tag{A.24}$$

where the complex relative dielectric function $\tilde{\varepsilon}_r$ is defined as

$$\tilde{\varepsilon}_r = \varepsilon_r + i \frac{\mu_0 \mu_r c^2 \sigma}{\omega} = \varepsilon_r + i \frac{\sigma}{\varepsilon_0 \omega} = \varepsilon_1 + i \varepsilon_2, \tag{A.25}$$

with its real and imaginary parts ε_1 and ε_2, respectively (note that ε_r is in general complex, thus $\varepsilon_r \neq \varepsilon_1$). The transformation $\varepsilon_r \to \tilde{\varepsilon}_r$ transforms the solution Eq. (A.24) of the inhomogenous wave equation Eq. (A.20) to look like a solution of the respective homogenous wave equation, which misses the second term on the right-hand side of Eq. (A.20). A homogenous wave equation has the general form

$$\nabla^2 \vec{\psi}(\vec{r}, t) = \frac{1}{u^2} \frac{\partial^2 \vec{\psi}(\vec{r}, t)}{\partial t^2} \tag{A.26}$$

where u denotes the propagation speed through the wave-carrying medium and in this case $u = \frac{1}{\sqrt{\varepsilon_0 \varepsilon_r \mu_0 \mu_r}}$. By introducing the complex-valued $\tilde{\varepsilon}_r$, we can conveniently write down a "complex-valued propagation speed" of the inhomogenous wave equation as

$$u = \frac{1}{\sqrt{\varepsilon_0 \tilde{\varepsilon}_r \mu_0 \mu_r}}. \tag{A.27}$$

Using the well-known relation for the speed of light in vacuum $c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$, we can write

$$u = \frac{c}{\sqrt{\tilde{\varepsilon}_r \mu_r}} = \frac{c}{\tilde{n}}, \tag{A.28}$$

and (with $\mu_r = 1$)

$$\tilde{n} = \sqrt{\tilde{\varepsilon}_r} = n + ik \tag{A.29}$$

where $\tilde{n}$ is known as the complex index of refraction. Its real part n denotes the factor by which the propagation speed of light inside the medium is reduced with respect to the vacuum and is called the index of refraction, and its imaginary part k is called the extinction coefficient, reducing the amplitude of the wave as it travels forward. Both can easily be verified by looking at the plane waves that emerge from Eq. (A.33) with the use of Eq. (A.24) and Eq. (A.29):

$$\vec{E}(\vec{r}, t) = \vec{E}_0 e^{i((\frac{\omega}{c}(n+ik))\vec{\kappa}_0 \cdot \vec{r} - \omega t)} = \vec{E}_0 e^{i\frac{\omega}{c} n \vec{\kappa}_0 \cdot \vec{r}} e^{-\frac{\omega}{c} k \vec{\kappa}_0 \cdot \vec{r}} e^{-i\omega t}, \tag{A.30}$$

where $\vec{\kappa}_0$ is a vector of unit length indicating the direction of propagation. The first exponential is is the oscillatory part and its exponent is proportional to $(c/n)^{-1}$, the second is a term decaying with kr.

In later sections we will make use of the relations

$$\varepsilon_1 = n^2 - k^2 \tag{A.31}$$

and

$$\varepsilon_2 = 2nk. \tag{A.32}$$

Since any superposition of plane waves of the form Eq. (A.33) solves the Macroscopic Maxwell Equations in the medium through the relation Eq. (A.24), the entire problem of the plethora of microscopic charges and currents in a solid ultimately reduces to finding the response function $\tilde{\varepsilon}_r$ or $\tilde{n}$ of the medium to light of all the frequencies, and finally stitching the so-found solutions together at the interfaces of different media. The rest of the sections in this chapter will therefore mainly focus on the properties of $\tilde{\varepsilon}_r$, and through it, $\tilde{n}$. This was a simple text book derivation, as can be found for instance in [161–163].

A.3 Dielectric Function

A.3.1 Refractivity, Dispersion, and Absorption

The macroscopic phenomena of *dispersion* and *absorption* of light can be seen as complementary processes, although they are fundamentally linked and just different spectral regimes of the same interaction. It will become clear, that dispersion is related to the refractive index n and absorption to the extinction coefficient k, the real and imaginary parts of the complex index of refraction given in Eq. (A.29). For structural purposes of this chapter, they will initially be seen as separate and dispersion will be discussed in the *transparent* regime of a medium first. This only implies that the interaction takes place far away from any absorption, not that the transparent region necessarily coincides with the visible range.

A.3.1.1 Refractivity

This content is not published online.

A.4 General Remarks

This content is not published online.

This content is not published online.

This content is not published online.

This content is not published online.

This content is not published online.

This content is not published online.

This content is not published online.

This content is not published online.

This content is not published online.

A.4.0.2 Dispersion (General)

The term optical dispersion denotes the frequency dependence of the phase velocity of an optical wave in a medium. This term originates from the well-known fact that a transparent prism, for instance made of glass and hit by white (visible) light, disperses the different colors of the rainbow into different directions. This is a consequence of the fact that the angle of refraction at a smooth surface between two media depends on the contrast of the phase velocity of light in these two media. Thus different phase velocities of the different colors lead to different angles of refraction.

The phase velocity is just c/n, where c is the speed of light in vacuum, and n is the "refractive index". It is often referred to as the "optical constant" of a medium, which is a statement of the fact that $n(\omega)$ normally varies only slightly with frequency in most of the spectral region, where a medium is transparent. This chapter discusses a simple phenomenological picture for the general shape of the *dispersion relation*, the more fitting term for $n(\omega)$. The basic formulation can be found in many textbooks, e.g. [161].

Consider atoms as driven harmonic oscillators that are made up of electrons that are elastically connected to much heavier nuclei. The dipole moments $\vec{p}$ (taken to be parallel to the x-axis) are then proportional to the deviation $\vec{x}$ from their central resting position, $\vec{p} = -e\vec{x}$. The electrons obey the equation of motion ($\vec{x} = x\vec{e}_x$):

$$m\ddot{x} + b\dot{x} + Dx = -eE_0 e^{i(\omega t - kz)}, \tag{A.33}$$

where $\vec{E} = E_0\vec{e}_x e^{i(\omega t - kz)}$ is the driving field (propagating along the z-direction and plane-polarized in the x-z plane), D is the "spring constant" and b the damping factor. The eigenfrequency is consequently $\omega_0^2 = D/m$ and the broadening term $\gamma = b/m$. With the Ansatz $x = x_0 e^{i(\omega t - kz)}$, one obtains the solution

$$x_0 = -\frac{eE_0/m}{(\omega_0^2 - \omega^2) - i\gamma\omega}. \tag{A.34}$$

If we set the the momentary phase of the driving field as the reference, and we consider a dipole at $z = 0$, then the phase shift of the dipole is just the argument of the complex quantity given in Eq. (A.34)

$$\tan\Delta\phi = -\frac{\gamma\omega}{\omega_0^2 - \omega^2}, \tag{A.35}$$

Interestingly, the fact that refractivity and dispersion of media are related to polarization

phenomena of charged particles, dates back to the 1870s, a time when Maxwell's theory of radiation was already known, but the concept of electrons was not. In 1871 Lord Rayleigh published his work on scattering using a generic polarizable atomic dipole and between 1872 and 1875 Sellmeier and Helmholtz explained how the coherent emission of dipoles in a solid that oscillate in phase with the driving field leads to the observable form of the dispersion relation $n(\omega)$ (far away from any absorption). This was consequently termed the "Mitschwingen model" (German, "to oscillate along"). The starting point of its derivation was the same as was done above, except for the fact that neither the nature of the oscillating particles, electrons, were known, nor was the nature of the phase shift derived explicitly, but rather postulated to be 90°, also, there was no damping term. This lead to the well-known Sellmeier Formula for the refractive index, which is still in use today to describe the dispersion of transparent media, far away from all eigenfrequencies, and is treated as a purely phenomenological model with empirically fitted parameters K_i and ω_i:

$$n^2 = \varepsilon_r = 1 + \sum_i \frac{K_i}{\omega_i^2 - \omega^2}. \tag{A.58}$$

Shortly after the discovery of the electron and its charge-to-mass ratio in 1896 by J.J. Thomson, in 1904, Paul Drude identified the oscillating particles with electrons bound to heavy nuclei and identified the weight factors of the Sellmeier Formula, K_i, with the relation $N_i e^2/\varepsilon_0 m$, where N_i was the number of electrons of one oscillator and e and m the charge and mass, respectively. His derivation was implicitly used here to arrive at Eqs. (A.55) and (A.56). To arrive at an expression for the index of refraction from these considerations, he took the following steps:

The polarization density of the medium was said to be $\vec{P} = N\vec{p}$, where N is the electron density. But $\vec{P}$ is also $\varepsilon_0 \chi_e \vec{E}$, where χ_e is the electric susceptibility of the (linear, isotropic) medium. With that it follows that

$$\varepsilon_0 \chi_e = \frac{Ne^2/m}{(\omega_0^2 - \omega^2) - i\gamma\omega} = \left(\frac{Ne^2}{m}\right) \frac{(\omega_0^2 - \omega^2) - i\gamma\omega}{(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2} \tag{A.59}$$

for a single oscillator. But since the relative permittivity $\varepsilon_r = \varepsilon_1 + i\varepsilon_2 = 1 + \chi_e$ (see Eq. (A.38)), where ε_1 and ε_2 are its real and imaginary parts, it follows that

$$\varepsilon_r = 1 + \frac{e^2}{\varepsilon_0 m} \sum_i \frac{N_i}{(\omega_i^2 - \omega^2) - i\gamma_i\omega} \tag{A.60}$$

and

$$\varepsilon_1 = 1 + \frac{e^2}{\varepsilon_0 m} \sum_i \frac{N_i(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + (\gamma_i\omega)^2} \tag{A.61}$$

$$\varepsilon_2 = \frac{e^2}{\varepsilon_0 m} \sum_i \frac{N_i\gamma\omega}{(\omega_0^2 - \omega^2)^2 + (\gamma_i\omega)^2}, \tag{A.62}$$

where the intuitive generalization has been made that there are sets of mutually exclusive oscillators i that have subscripted quantities N_i, ω_i, and γ_i and their overall polarization

density simply adds up.

Shortly after this derivation, Drude even made the bold move to suggest that these so-called "dispersion electrons" where to be identified with "valence electrons", in accordance with recent works connecting electron theory to chemistry, and thereby making a connection between the new field of molecular chemistry and optics, thereby suggesting the use of optical methods as tools for investigating the periodic table.

Equations (A.60)-(A.62) are called the Drude-Lorentz-Oscillator Model and its consequences of are discussed next. ε_r is a smooth, real function except near the resonant frequencies ω_i. There, the imaginary part, ε_2 sharply increases. At these points the magnitude of the polarization $\propto \sqrt{\chi_e}$ is maximum and the re-emission $\propto |\vec{\vec{P}}|$ is 180° out of phase with the driving field. This is due to a 90° contribution from the dipole oscillation itself and 90° retardation from the reemission, whose amplitude relative to the driving field is large at the resonance due to the large polarizability there. This leads to destructive interference in the forward direction and prohibits wave propagation, which becomes clear when remembering Eq. (A.32) that states that $\varepsilon_2 = 2nk$ and k introduces an exponential decay of the overall light field in the forward direction.

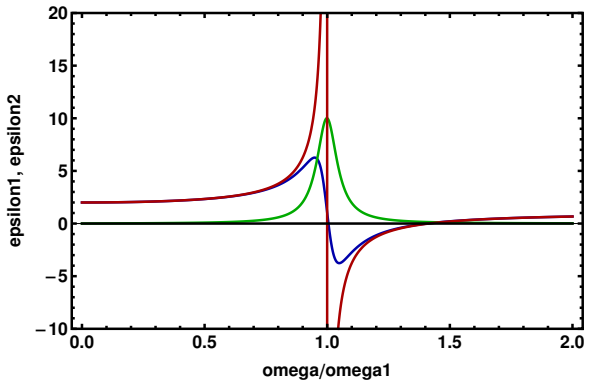


Figure A.2 *Plot of real and imaginary parts of the dielectric function, ε_1 and ε_2, respectively, for a solid modeled with a single damped (undamped) oscillator according to the Drude-Lorentz Oscillator model. Blue line: ε_1*(damped)*; Green line: ε_2*(damped)*; Red line: ε_1*(undamped)*; Black line ($\equiv 0$): ε_2*(undamped)*. Note that according to Eq.(A.62) the imaginary part is proportional to the damping, therefore the undamped oscillator's dielectric function is purely real.*

Figure A.2 is a plot of Eqs. (A.61) and (A.62) as a function of frequency for a solid modeled with a single oscillator, whereby situations for a damped and undamped oscillator are shown. There is a sharp oscillation of ε_1 in the vicinity of the resonance frequency, which for the undamped oscillator turns out to be a singularity. The imaginary part ε_2 of the undamped oscillator exhibits sharp resonance behavior around the resonance frequency, whereas the undamped oscillator is a constant zero. Figure A.3 is a plot of the real and imaginary parts, n and k, of the square root of the full dielectric function, Eq. (A.60), as a function of frequency for the same solid, in the damped and undamped cases. In the undamped case, n goes to infinity at the resonance, to then sharply drop off to zero and to then steadily increase again from zero at a higher frequency. In the range where n is zero, k has values different to zero, starting off at infinity at the resonance and then quickly dropping to zero again at the point

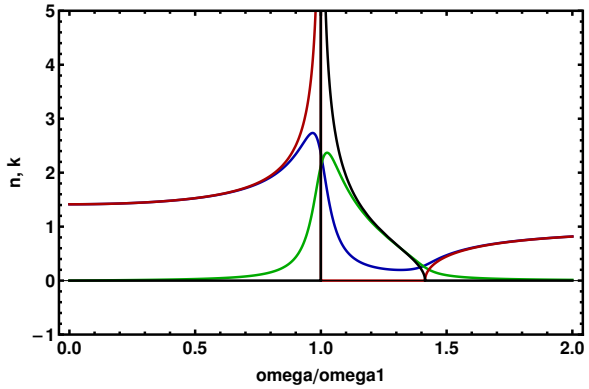


Figure A.3 *Plot of refractive index and extinction coefficient, n and k, respectively, for a solid modeled with a single damped (undamped) oscillator according to the Drude-Lorentz Oscillator model. Blue line: n*(damped)*; Green line: k*(damped)*; Red line: n*(undamped)*; Black line: k*(undamped)*. Note that since the undamped oscillator's dielectric function is purely real, $n \equiv 0$ in ranges where k is finite and vice-versa.*

where n has finite values. Thus in the undamped case, the solid has finite values for n and k in mutually exclusive ranges. In the damped case, the sharp peak is rounded off and n and k mix in the same ranges. Note that $n(0) \neq n(\infty) = 0$ and that $n < 1$ above the resonance, thus the phase velocity is actually *faster* than the speed of light in vacuum. This explains why X-rays propagate faster in most solids than they do in the vacuum. Also note that *except* in a narrow range around the resonance for the damped case, $n(\omega(A)) < n(\omega(B))$ if $\omega(A) < \omega(B)$. This is called normal dispersion. If the contrary is true, it is called anomalous dispersion. It is also important to note that except in a narrow range around the resonance frequency, say between 0.8 and 1.2, ε_1(damped) and ε_1(undamped) are virtually identical, and so are n and k in their damped and undamped versions. Thus not so far from the resonances, undamped oscillators are a good substitute for the more complicated damped versions, which is one of the underlying reasons for the success of the Sellmeier Model.

The simple picture of "electrons on springs" of the Drude-Lorentz Model is of course not to be taken too seriously or overinterpreted. The real use of the model is to be able to fit complicated experimental data to relatively few open parameters with reasonable fidelity. However, it is instructive to mention one important correction to the form of Eq. (A.60). During the discussion in conjunction with the Macroscopic Maxwell Equations Eqs. (A.9) – (A.12), it was mentioned that all the microscopic electric fields average out on a macroscopic level and result in a macroscopic polarization density, whose field overlies the externally applied electric field. This is only true for the macroscopic average though. On a microscopic level, it is clear that there must be a strong oscillation of the fields close to the atomic sites. The superposition of these microscopic fields at the position of the individual dipoles, $\vec{E}^{\text{local}}$, is what actually drives the oscillations in Eq. (A.55). It turns out that this local field can be rather well modeled as a superposition of four terms

$$\vec{E}^{\text{local}} = \vec{E}^{\text{ext}} + \vec{E}^{\text{pol}} + \vec{E}^{\text{L}} + \vec{E}^{\text{int}} = \vec{E}^{\text{diel}} + \vec{E}^{\text{L}} + \vec{E}^{\text{int}}, \tag{A.63}$$

where $\vec{E}^{\text{ext}}$ is the externally applied field (the impinging light) and $\vec{E}^{\text{pol}} = -\vec{P}/\varepsilon_0$ is due to the macroscopic polarization density. The so-called Lorentz field $\vec{E}^{\text{L}} = \frac{\vec{P}}{3\varepsilon_0}$ is the contribution of the polarization of a virtual mesoscopic sphere centered at the atomic dipole in question and $\vec{E}^{\text{int}}$ is the sum of the contributions from all microscopic dipoles inside the mesoscopic Lorentz sphere. Therefore $\vec{E}^{\text{int}}$ actually depends on the crystal structure of the solid and $\vec{E}^{\text{int}} = 0$ for monoatomic cubic lattices and amorphous solids. Implicitly, it was $\vec{E}^{\text{diel}} = \vec{E}^{\text{ext}} + \vec{E}^{\text{pol}}$ that was used as the driving force in Eq. (A.55). The corrected field to use is

$$\vec{E}^{\text{local}} = \vec{E}^{\text{diel}} + \frac{\vec{P}}{3\varepsilon_0} = (1 + \chi/3)\vec{E}^{\text{diel}}, \tag{A.64}$$

substituting $\vec{P} = \chi_e \varepsilon_0 \vec{E}^{\text{diel}}$. This leads to formally the same expression as Eq. (A.60), except that the resonance frequencies ω_i become shifted to

$$\bar{\omega}_i = \sqrt{\omega_i^2 - \frac{Ne^2}{3\varepsilon_0 m}}. \tag{A.65}$$

The Drude-Lorentz-Oscillator Model of optical dispersion forms the basis of most modern semi-empirical dispersion relations, including the Brendel Oscillator Model (Appendix B.2.3), the Tauc-Lorentz Model [192], the Cody-Lorentz Model [192], and in its simplest form without the damping term, the Sellmeier Model. Its importance stems from the fact that in the linear optics regime, where wavelengths are large compared with the granularity of the solid, homogeneous, non-ferromagnetic materials can in fact be modeled correctly with the use of phenomenological classical harmonic oscillators appropriately distributed over the frequency spectrum. This becomes evident, when the full quantum mechanical treatment of the microscopic interactions results in expressions that are structurally identical to Eq. (A.60), and allow for identification of the classical quantities that define each oscillator with physical observables.

A.4.0.3 Absorption

Electromagnetic waves traveling through a sample can be absorbed to change the internal quantum state of the sample. The sample thereby removes power from the wave traveling into the forward direction. This is called a *non-parametric process*. For a sample to be excited to a higher energy state by incident light, a number of conditions have to be met. First, there has to exist an energy level spaced exactly as far (upward) from the occupied level as is associated with the frequency of the light. Second, this excited state has to be unoccupied. Third, the states have to couple through an interaction that involves the light. Fourth, this transition cannot be forbidden by symmetry or other selection rules. And last, if we view the sample as a set of different interlinked quantum systems (e.g. electrons, nuclei, lattice vibrations), the more final states there are within a small energy range that corresponds to the energy of the light, the stronger the interaction, and thus the absorption, will be. Therefore absorption experiments are a vital tool to probe the internal quantum structure of the sample. The basic math can be found, for instance, in [161].

In absorption measurements, the absorption coefficient is accessible through the following formula

$$I(z) = I(0)e^{-\alpha_{\text{abs}}(\omega)z}, \tag{A.66}$$

where $I(z)$ is the intensity of the light at position z as it travels through the medium and α_{abs} is the absorption coefficient. This relation simply follows from Eq. (A.30) (for a wave traveling in the z-direction), considering that $I(z) \propto |\vec{E}(z)|^2$ and defining

$$\alpha_{\text{abs}} = 2\frac{\omega k(\omega)}{c}. \tag{A.67}$$

When evaluating measurements according to Eq. (A.66), it is important to consider the Fresnel reflections R that occur at every interface, and in certain cases can even lead to multiple internal reflections. In the limit of thick samples and high absorption, where multiple internal reflections can be neglected, the proper form for the initial intensity is $I(0) = I_0(1 - R)$, where I_0 is the total intensity of the light shone on the sample, and the transmitted intensity is

$$I_{\text{transmitted}} = I_0(1 - R)^2 e^{-\alpha_{\text{abs}}(\omega)L}, \tag{A.68}$$

where L is the thickness of the sample. In any case, due to conservation of energy, transmitted power, reflected power, and absorbed power have to balance out:

$$T + R + A = 1. \tag{A.69}$$

One look at Fig. A.3 reveals that absorption will only take place in the vicinity of a resonance frequency, since $\alpha_{\text{abs}} \propto k$.

A matter of considerable debate in the early days of quantum mechanics was what exactly constitutes the nature of the resonances in the Drude-Lorentz Oscillator Model, until it was discovered that the energies associated with the frequencies are in fact identical to the transition energies of different quantum states of the medium (solid, liquid, gas). Today it is clear, that the Drude-Lorentz Oscillators can be derived very elegantly when treating the medium as a quantum mechanical system perturbed by electromagnetic radiation. In second-order perturbation theory, the transition rates between states that can couple through a perturbing radiation Hamiltonian show resonant behavior at frequencies associated with the spacing of the energy levels in a shape that looks exactly like Drude-Lorentz Model. In Eq. (A.60), the resonance frequencies ω_i become $\frac{E_2 - E_1}{\hbar}$ where E_1 and E_2 are the energies of the initial and final states and $\hbar$ is the reduced Planck constant. The weight factors N_i become a product of the Joint Density of States (JDOS) for the medium $\rho(\omega_i)$ and the transition matrix elements $\langle 1| \mathcal{H}' |2\rangle$ between the initial and final states (coupled through the radiation Hamiltonian $\mathcal{H}'$), which for a gas are also proportional to the Einstein B coefficients. The damping terms γ_i result from the finite life times of the excited states, which can relax into any of the lower states through radiative and non-radiative processes. For gases, the radiative relaxations are

related to the Einstein A coefficients. The sum considers which states are occupied and unoccupied, if an occupied state cannot be occupied twice, as for electrons. For very densely lying states the sum is approximated by a continuous function using the aforementioned Joint Density of States (JDOS) as the frequency weight. The full quantum-mechanical treatment will be done in some of the following sections where necessary, as for Interband Absorption.

It is interesting to take another closer look at the qualitative behavior of the Drude-Lorentz Model. Although the term "absorption coefficient" suggests that the interaction causes light to be lost to other channels such as heat, this is not necessarily the case. α_{abs} is proportional to the extinction coefficient k, which only states that light is hindered in *propagating*. Yet, the stronger the extinction coefficient k, the greater also the reflection at an interface with the medium, which is clear from the fact that the (intensity) reflectivity is

$$R(\omega) = \frac{(1 - n(\omega))^2 + k(\omega)^2}{(1 + n(\omega))^2 + k(\omega)^2} \tag{A.70}$$

at normal incidence, which follows from Fresnel's Equations. The Reststrahlen phenomenon is an illustrations of this. The Reststrahlen band is the name for the narrow frequency band just above a strong phonon interaction (discussed in a later section), where nearly total reflection of incident radiation occurs. The transverse-optical phonon absorption looks surprisingly similar to a nearly undamped single Drude-Lorentz oscillator as pictured in Fig. A.3. In the range just above the resonance, k is large and n is small, giving rise to strong reflection. This is a statement of the fact that the involved absorption and emission processes leading to a large k are largely elastic. It can in fact be shown, that actual *loss* of radiation is proportional to the imaginary part of the dielectric function ε_2, which is proportional to a product of n and k through Eq. (A.32). Which is proportional to the *damping* of the oscillator and is therefore smaller, the less damped the interaction is. This fact normally gets ignored in text books, since the Restrahlen band is the the only good example for a situation where the absortion (through the extinction k) is large, but the loss (through ε_2) is small. In metals below the plasma frequency (discussed in a later section) and in interband absorption, n and k are large at the same time and so is the loss.

It is still highly instructive to make the following distinction: The real part of the dielectric function ε_1 is a measure for how well a material can *store* energy from electromagnetic radiation through its polarization. The imaginary part ε_2 is a measure for how well a material can *dissipate* energy from electromagnetic radiation through non-radiative relaxation channels. The index of refraction n is a measure for how *slowly light travels* through a medium, and the extinction coefficient k is a measure for how *well light propagates* through it.

Finally, one brief comment on the time scale of dispersive and absorptive interactions. Absorption of a photon by some quantum system – an atom, molecule, or solid – leads to a change of the internal quantum state of the system, and with it its internal energy. This photon can now be either lost to internal non-radiative interactions, or it can be re-emitted at the same frequency or a different one, elastically or inelastically, respectively. The state change of the absorbing system can be observed, and the timescale of this process is governed

by things like coupling matrix elements, selection rules, and densities of states, ultimately determining a probabilistic lifetime of the excited state. These processes are called *non-parametric*. On the other hand, as mentioned earlier in this section in conjunction with dispersion (and coherent Rayleigh scattering), the quantum mechanical description of the scattering process also lets the atom "absorb" a photon out of the light field and re-emit it. In this case though, the polarization of the atom does not equal an electronic excitation. Light scattering can take place at arbitrarily low photon energies at spherically symmetric isolated atoms (to avoid the option of rotations or vibrations), even if the nearest unoccupied electronic state is energetically far away. Thus the internal energy of the atom has no physical means of changing. The traditional explanation of this is that the polarization of the atom upon absorption of a photon equals a virtual state with no actual energy change, thus – to not violate energy conservation – the process is inherently "instantaneous". Processes of this sort are called *parametric*.

A.4.1 Kramers-Kronig Relations and Sum-Rules

From measurements of the absorption coefficient the extinction coefficient k can be determined, whereas the reflectivity of a sample's surface is determined by a complicated mix of the index of refraction n and the extinction coefficient. Thus often there is insufficient information to determine ε_1 and ε_2 independently. However, if either n or k, or ε_1 ε_2 are known over a wide frequency range, the other can be computed through the Kramers-Kronig relations [161]:

$$\varepsilon_1(\omega) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{\omega'^2 - \omega^2} d\omega', \tag{A.71}$$

$$\varepsilon_2(\omega) = -\frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon_1(\omega')}{\omega'^2 - \omega^2} d\omega' \tag{A.72}$$

and

$$n(\omega) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\omega' k(\omega')}{\omega'^2 - \omega^2} d\omega', \tag{A.73}$$

$$k(\omega) = -\frac{2\omega}{\pi} P \int_0^\infty \frac{n(\omega') - 1}{\omega'^2 - \omega^2} d\omega', \tag{A.74}$$

where P denotes the principal value of the integral. These relations are based on causality, linear response theory and the boundedness of physical observables. Their derivation will not be explicitly executed here, but can be found in any text book on the subject.

Outside of being very useful when calculating dielectric functions, they can be used to make a few very general and interesting statements about the response of materials to electromagnetic radiation and the self-consistency of the results.

First of all, they explain why this chapter on dielectric properties goes through the trouble of identifying interactions as far as the ultra-violet range, when this work is merely on infrared

materials. For this, let us consider the (quasi)-static dielectric constant of a solid without free carriers ($\varepsilon_2(0) = 0$) as given by Eq. (A.71),

$$\varepsilon_1(0) = 1 + \frac{2}{\pi}\int_0^\infty \frac{\varepsilon_2(\omega')}{\omega'}d\omega'. \tag{A.75}$$

It is clear that all absorption peaks of ε_2 contribute, but also they are weighted less the higher they lie in frequency and the more, the broader and stronger they are. This can be illustrated quite well with the first solids of group IV of the periodic table. Diamond, silicon, and germanium are all in the diamond structure, all covalent, no infrared-active phonons, nearly no free carriers, and generally quite similar in structure. Their interband transition spectra look relatively similar as will be shown in a later section. Their band gap is vey different though, thus the onset of the absorption spectrum happens higher and higher in the order Ge, Si, Dia. Thus from Eq. (A.75) we expect diamond to have the lowest refractive index at low frequencies and germanium the highest. This is very much the case: Diamond with a bandgap of 5.5eV has $n = 2.4$ and $\varepsilon_1 = 5.8$, silicon with a bandgap of 1.1eV has $n = 3.4$ and $\varepsilon_1 = 11.7$, and germanium with a bandgap of 0.7eV has $n = 4.0$ and $\varepsilon_1 = 16.0$. Thus the index of refraction in the infrared range is solely determined by far-lying interband transitions.

One very general remark is, that if we want to know the refractive index or ε_1 in a certain energy range, then it is most important to know k or ε_2 in and close to that energy range, because through the denominators the contribution of that range is weighted most strongly. More often than not some sort of extrapolations have to be made though, which can be useful, but can also lead to significant errors in the calculated functions.

Another interesting thing to point out is the generalization of the f-sum rule as it applies to the classical Drude-Lorentz Model, Eq. (A.60). Equation (A.60) can be rewritten in the form

$$\varepsilon_r = 1 + \omega_p^2 \sum_i \frac{f_i}{(\omega_i^2 - \omega^2) - i\gamma_i\omega}, \tag{A.76}$$

where $\omega_p = \sqrt{\frac{Ne^2}{\varepsilon_0 m}}$ is the plasma frequency of all contributing electrons, $N = N'Z$, where N' is the number density of the atoms and Z the number of electrons per atom. f_i is known as the *oscillator strength* of the different contributions. The f-sum rule states that

$$\sum_i f_i = 1. \tag{A.77}$$

The generalization that can be derived from the Kramers-Kronig-Relations is

$$\int_0^\infty \omega'\varepsilon_2(\omega')d\omega' = \frac{\pi}{2}\omega_p^2. \tag{A.78}$$

A consequence of Eq. (A.79) is that an "effective" number of electrons can be computed that participate in oscillations up to a given frequency [187]:

$$N_{\text{eff}}(\omega) = \frac{2m\varepsilon_0}{\pi e^2} \int_0^{\omega} \omega' \varepsilon_2(\omega') d\omega'. \tag{A.79}$$

Other interesting sum rules include the inertial sum rule [187]

$$\int_0^{\infty} (n(\omega) - 1) d\omega = 0 \tag{A.80}$$

and the DC conductivity sum rule

$$\int_0^{\infty} (\varepsilon_1(\omega) - 1) d\omega = -2\pi^2 \sigma_0, \tag{A.81}$$

where σ_0 is the DC electrical conductivity as given in the next chapter.

A.4.2 Drude Free-Carrier Contribution

Every charge that is not bound to a certain position inside a solid, but can move freely within the boundaries of the solid itself, is termed "free". In metals, free electrons exist even at the lowest of temperatures, whereas in semiconductors, carriers need finite thermal energy to become untied. In semiconductors, the number and type of carriers, electrons or holes, can be controlled through the use of dopants.

Free carriers are surprisingly well modeled using the Drude Model that treats them as an ideal gas of non-interacting particles that only scatter at the stationary ionic lattice with a phenomenological scattering time. The well-known Drude conductivity is given by [161]:

$$\sigma^{\text{Drude}}(\omega) = \frac{ne^2\tau}{m^*(1 - i\omega\tau)}, \tag{A.82}$$

where n is the number density of free carriers, e is the elementary charge (of a carrier), τ the scattering time, and m^* the effective mass. The conductivity is oblivious to the type of carrier, electron or hole, since the sign of the charge is lost due to the square.

The conductivity enters the dielectric function through Eq. (A.25), which is to be read as

$$\varepsilon_r = \varepsilon_r^{\text{other processes}} + \frac{i}{\varepsilon_0 \omega} \sigma^{\text{Drude}}, \tag{A.83}$$

where $\varepsilon_r^{\text{other processes}}$ refers to all other contributions to the dielectric function, particularly all the contributions of the bound ion cores. The tilde of $\tilde{\varepsilon}_r$ is omitted in this section for simplicity.

At very low frequencies, $\omega \to 0$, a condition that for metals is normally still well met up to the visible range, it is clear from Eq. (A.83), that the imaginary part of the dielectric function diverges, $\varepsilon_2 \to \infty$, while the real part tends towards the constant $\varepsilon_1 \to -\frac{ne^2\tau^2}{\varepsilon_0 m^*}$, which for metals is normally a very large number. A dielectric function of diverging magnitude is to be understood as follows. Since the charges are not bound, at sufficiently low frequencies, they can freely follow external driving fields over arbitrary distances leaving bare the immobile ions of the opposite charge. Thus, given a sufficiently large number of free carriers for a

given field intensity, the inside of a solid is completely shielded from electromagnetic fields by carrier response on its surface. In other words, the electric susceptibility is arbitrarily large as the frequency tends to zero, thus sustaining a finite overall polarization of the solid, with zero internal electric field. The large real part is a statement for the fact that the solid stores energy well at low frequencies, due to the large sustainable fields over macroscopic distances that are able to cancel any external field fully. The diverging imaginary part is a statement of the fact that the conductivity dissipates the stored energy very well into heat. Thus, in this regime of very strong free-carrier absorption, where $|\varepsilon_r^{\text{other processes}}| \ll |\frac{\sigma}{\varepsilon_0 \omega}|$, we obtain a nearly purely imaginary dielectric function

$$\varepsilon_r \approx i\frac{\sigma}{\varepsilon_0 \omega} \approx i\frac{ne^2\tau}{\omega m^*} \approx i\varepsilon_2 = 2ink, \tag{A.84}$$

and since the square root of an imaginary number has equally large real and imaginary parts, it follows that $n \approx k$ and both diverge toward $+\infty$ for small frequencies.

The quantity

$$\omega_p = \sqrt{\frac{ne^2}{m^*\varepsilon_r\varepsilon_0}} \tag{A.85}$$

is known as the plasma frequency and is an approximate expression of the point where the real part of the dielectric constant turns zero, $\varepsilon_1(\omega_p) = 0$.

To interpret the plasma frequency, it is instructive to see how the dielectric function behaves in the vicinity of the plasma frequency. For metals, where the concentration of free carriers n is huge, it is generally true that $\omega_p\tau \gg 1$. In this case,

$$\varepsilon_r \approx \varepsilon_r^{\text{other processes}} \left(1 - \frac{\omega_p^2}{\omega^2} + i\frac{\omega_p^2}{\omega^3\tau}\right) \tag{A.86}$$

where the third term in brackets is substantially smaller than the other two because in the vicinity of ω_p, obviously $\omega\tau \gg 1$. If we make the assumption that $\varepsilon_r^{\text{other processes}}$ is real and positive, which is also generally true far away from all other interactions, $\varepsilon_r < 0$ and almost purely real for $\omega \lesssim \omega_p$ and $\varepsilon_r > 0$ and almost purely real for $\omega \gtrsim \omega_p$. Since according to Eq. (A.29), $\tilde{n} = n + ik = \sqrt{\varepsilon_r}$, the complex index of refraction is either purely imaginary for $\omega < \omega_p$ or purely real for $\omega > \omega_p$. Thus the plasma frequency is the point where a metal becomes transparent to electromagnetic radiation. It can be shown that the plasma frequency is the resonance frequency of the free electrons around the bound ion cores, i.e. the frequency where the electrons can sustain oscillations. For metals the plasma frequency normally lies in the ultra-violet, hence the total reflection at visible wavelengths.

In the range of optical measurements in semiconductors, especially in the mid-infrared, the condition $\omega \gg \omega_p$ is generally satisfied due to the low density of free carriers and consequently low plasma frequency. It can be shown that the condition $\omega_p\tau \gg 1$ is also normally met above the optical phonon frequencies (to be discussed in a later section), thus the approximate expression Eq. (A.86) holds and reduces to $\varepsilon_r \approx \varepsilon_r^{\text{other processes}}$. Therefore in a semiconductor at all but the lowest of frequencies the free-carrier contribution can be neglected.

Finally, it is important to point out that the free carrier absorption can be modeled with a Drude-Lorentz Oscillator as given by Eq. (A.60), by setting the resonance frequency $\omega_i = 0$ and the damping constant $\gamma_i = 1/\tau$. The number of electrons in the oscillator N_i is then just the number of free carriers.

A.4.3 Interband Transitions

Interband transitions arise when a photon has enough energy to make an electron from an occupied state in one electronic band jump to an unoccupied state in another band. These transitions are therefore to be distinguished from intra-band transitions which can only occur in bands that are partially filled and therefore make up the (Drude) free-carrier contribution discussed before. The full dielectric function ε now has the form [161]:

$$\varepsilon = \varepsilon^{\text{other processes}} + \frac{i}{\varepsilon_0 \omega}(\sigma^{\text{Drude}} + \sigma^{\text{interband}}), \tag{A.87}$$

where all other processes that are not explicitly regarded here are compounded in $\varepsilon^{\text{other processes}}$, and the free-carrier contribution to the conductivity is σ^{Drude}. $\sigma^{\text{interband}}$ is the conductivity tensor arising from the interband contribution, whose components (α, β) are defined by

$$j_\alpha^{\text{interband}} = \sigma_{\alpha\beta}^{\text{interband}} E_\beta, \tag{A.88}$$

where the current density $j^{\text{interband}}$ is the interband contribution to the full current density, which is the only contribution if the electric field E has a frequency high enough for free-carrier contributions no longer to play a role. This is of course not the case for metals. The components then take on the form:

$$\sigma_{\alpha\beta}(\omega) = -\frac{e^2}{m^2} \sum_{i,j} \frac{|f(E_i) - f(E_j)|}{E_i - E_j} \frac{\langle i|\, p_\alpha \,|j\rangle \langle j|\, p_\beta \,|i\rangle}{-i\omega + 1/\tau + (i/\hbar)(E_i - E_j)}, \tag{A.89}$$

which can be inferred through linear response theory using a Green's function method, or Fermi's Golden Rule, but will not be done here explicitly. Instead, the parts making up Eq. (A.89) will be discussed in the following.

- The contribution is a sum over all initial states i of the valence band and the final states j of the conduction band.
- The Fermi functions $f(E_i)$ and $f(E_j)$ ensure that only those states contribute to the initial states that are occupied and those to the final states that are unoccupied (spontaneous emission has a negative contribution).
- Structure of the conductivity arises from the resonant nature of the denominator. Thus one particular transition will only contribute to the overall conductivity near a photonic energy $\hbar\omega$ that is equal to the energy separation of the two states $E_i - E_j$. The broadening around this energy is due to the relaxation time τ. A consequence of this is that the conductivity at $\hbar\omega$ is proportional to the joint density of states (JDOS) at

$\hbar\omega$, which is defined to be the number of different initial and final states that have an energy separation of $\hbar\omega$. This can be understood from the fact that the sum in Eq. (A.89) counts *all* transitions, and the contribution at $\hbar\omega$ will be large if there are many such transitions available.

- The transition matrix elements coupling the initial and final states are given by the components of the momentum operator $p_{\alpha,\beta}$ which gives the conductivity its tensorial character. The reason for the momentum operator's matrix elements to appear in this context will be discussed next.

The Hamiltonian for an electron in an optical field has the form

$$\mathcal{H} = \frac{1}{2m}(\vec{p} - \frac{e}{c}\vec{A})^2 + e\phi, \tag{A.90}$$

where $-e$ and m is the charge and mass of the electron, c is the speed of light, and ϕ and A are the scalar and vector potentials of the electromagnetic field. That this is true can be shown by using Hamilton's equations

$$\vec{v} = \frac{\partial \mathcal{H}}{\partial \vec{p}} = \frac{1}{m}(\vec{p} - \frac{e}{c}\vec{A}) \tag{A.91}$$

$$\dot{\vec{p}} = -\vec{\nabla}\mathcal{H} = -e\vec{\nabla}\phi + \frac{e}{c}\vec{\nabla}(\vec{A}\cdot\vec{v}) \tag{A.92}$$

to satisfy the classical equation of motion

$$\frac{d}{dt}(m\vec{v}) = e(\vec{E} + \frac{1}{c}(\vec{v}\times\vec{H})), \tag{A.93}$$

where $\vec{v}$ is the velocity of the electron. The electric field $\vec{E}$ has the form

$$\vec{E} = -\vec{\nabla}\phi - \frac{1}{c}\frac{\partial\vec{A}}{\partial t} \tag{A.94}$$

and the magnetic flux $\vec{B} = \mu\vec{H}$ (μ is the magnetic permeability and H the magnetic field) has the form

$$\vec{B} = \vec{\nabla}\times\vec{A}. \tag{A.95}$$

With this, this equation of motion Eq. (A.93) has the form

$$\frac{d}{dt}(m\vec{v}) = e(-\vec{\nabla}\phi - \frac{1}{c}\frac{\partial\vec{A}}{\partial t} + \frac{1}{c}(\vec{v}\times(\vec{\nabla}\times\vec{A}))). \tag{A.96}$$

Using standard vector identities, Eq. (A.96) becomes

$$\frac{d}{dt}(m\vec{v} + \frac{e}{c}\vec{A}) = \vec{\nabla}(-e\phi) + \frac{e}{c}\vec{\nabla}(\vec{A}\times\vec{v}), \tag{A.97}$$

which is obviously satisfied using Hamilton's equations, Eqs. (A.91) and (A.92). Thus, it

is clear that an optical field is introduced into an existing Hamiltonian simply by making the replacement $\vec{p} \mapsto (\vec{p} - \frac{e}{c}\vec{A})$.

When introducing an optical field to a solid, the full one-electron Hamiltonian becomes

$$\mathcal{H} = \frac{1}{2m}(\vec{p} - \frac{e}{c}\vec{A})^2 + V(\vec{r}) = \frac{p^2}{2m} + V(\vec{r}) - \frac{e}{mc}\vec{A}\cdot\vec{p} + \frac{e^2\vec{A}^2}{2mc^2}, \tag{A.98}$$

where $V(\vec{r})$ is the full periodic potential of the ions. Since optical fields in the regime considered in this work are weak, the Hamiltonian is just the Hamiltonian without the optical field with a perturbation of the form

$$\mathcal{H}' = -\frac{e}{mc}\vec{A}\cdot\vec{p} + o(A^2). \tag{A.99}$$

Thus the matrix elements used to treat the solid with an optical perturbation are

$$\langle i|\,\mathcal{H}'\,|j\rangle = -\frac{e}{mc}\vec{A}\cdot\langle i|\,\vec{p}\,|j\rangle\,. \tag{A.100}$$

Eq. (A.100) makes clear that the relevant transition matrix elements to treat the effect of light on a solid and that are expected to appear in the conductivity Eq. (A.89) are the matrix elements of the momentum operator $\vec{p}$. At this point it is useful to note the following operator relation

$$\vec{p} = \frac{m}{i\hbar}\left[\vec{r}, \mathcal{H}_0\right]_-\,, \tag{A.101}$$

which leads to

$$\langle i|\,\vec{p}\,|j\rangle = \frac{E_j - E_i}{i\hbar} m\,\langle i|\,\vec{r}\,|j\rangle = -i\omega m\,\langle i|\,\vec{r}\,|j\rangle\,, \tag{A.102}$$

where $\vec{r}$ is the dipole moment operator.

It is further interesting to see how the momentum matrix elements are related to the effective mass tensor $\frac{1}{\mathbf{m}_n^*}$ (of semi-classical electron transport) allowing optical measurements to be used to measure properties of band structure, and allowing band structure to be inferred to make predictions about the outcome of optical measurements.

For a given electronic dispersion relation $E_n(\vec{k})$ where $\vec{k}$ is a reciprocal lattice vector and n a band index, the effective mass is defined as

$$\left(\frac{1}{\mathbf{m}_n^*}\right)_{\alpha,\beta} = \pm\frac{1}{\hbar^2}\frac{\partial^2 E_n(\vec{k})}{\partial k_\alpha \partial k_\beta}. \tag{A.103}$$

This definition transforms the dispersion relation of a real solid to that of a free electron in the Sommerfeld model. This has the great advantage that it simplifies calculating transport properties, by making the semi-classical equations of motion look like the familiar free-electron form. The semi-classical equations of motion are

$$\dot{\vec{r}} = \vec{v}_n(\vec{k}) = \frac{1}{\hbar}\frac{\partial E_n(\vec{k})}{\partial \vec{k}} \tag{A.104}$$

$$\hbar\dot{\vec{k}} = -e\left[\vec{E}(\vec{r},t) + \frac{1}{c}\vec{v}_n(\vec{k}) \times \vec{H}(\vec{r},t)\right]. \tag{A.105}$$

Using Eqs. (A.104) and (A.103) leads to

$$\vec{a} = \frac{d\vec{v}}{dt} = \pm\left(\frac{1}{\mathbf{m}_n^*}\right)(\vec{k})\hbar\dot{k}, \tag{A.106}$$

which in turn leads Eq. (A.105) to become

$$\mathbf{m}_n^*(\vec{k})\vec{a} = \pm e\left(\vec{E} + \frac{1}{c}\vec{v}(\vec{k}) \times \vec{H}\right), \tag{A.107}$$

which looks like the free-electron Lorentz force, where $\vec{E}$ and $\vec{H}$ are only the externally applied fields and all the effects of the periodic potential of the solid are fully accounted for by making the transcriptions $\vec{p} \mapsto \hbar\vec{k}$ and $m \mapsto \mathbf{m}_n^*(\vec{k})$.

However, if the dispersion relation is explicitly calculated using one of the many methods available for calculating band structure, $\mathbf{m}^*$ can explicitly be related to either experimental observables in semi-empirical theories or other physical quantities in ab-initio calculations.

A convenient starting point for calculating band structure is $\vec{k} \cdot \vec{p}$ ("k dot p") theory. This formalism is a perturbation approach, valid in its lowest order for small wave vectors $\vec{k} \approx 0$ near the Γ point. The starting point are the full Bloch functions

$$\psi_{n,\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}u_{n,\vec{k}}(\vec{r}), \tag{A.108}$$

where $u_{n,\vec{k}}(\vec{r}) = u_{n,\vec{k}}(\vec{r}+\vec{R})$ is a periodic function in real space, $\vec{R}$ is a lattice vector of the underlying Bravais lattice, and n is the band index.The Bloch functions are inserted into the Schrödinger Equation of the full periodic potential

$$\mathcal{H}\psi_{n,\vec{k}}(\vec{r}) = E_n(\vec{k})\psi_{n,\vec{k}}(\vec{r}) = \left[\frac{p^2}{2m} + V(\vec{r})\right]e^{i\vec{k}\cdot\vec{r}}u_{n,\vec{k}}(\vec{r}) = E_n(\vec{k})e^{i\vec{k}\cdot\vec{r}}u_{n,\vec{k}}(\vec{r}) \tag{A.109}$$

with the momentum operator $\vec{p} = \frac{\hbar}{i}\vec{\nabla}$ acting on the Bloch functions as

$$\vec{p}e^{i\vec{k}\cdot\vec{r}}u_{n,\vec{k}}(\vec{r}) = e^{i\vec{k}\cdot\vec{r}}(\vec{p} + \hbar\vec{k})u_{n,\vec{k}}(\vec{r}). \tag{A.110}$$

Then the Schrödinger Equation, with some of its terms rearranged to be in the appropriate shape for a perturbation approach, takes on the form

$$\big[\underbrace{\frac{p^2}{2m} + V(\vec{r})}_{\mathcal{H}_0} + \underbrace{\frac{\hbar\vec{k}\cdot\vec{p}}{m}}_{\mathcal{H}'}\big]u_{n,\vec{k}}(\vec{r}) = \big[\underbrace{E_n(\vec{k}) - \frac{\hbar^2\vec{k}^2}{2m}}_{\hat{E}_n(\vec{k})}\big]u_{n,\vec{k}}(\vec{r}), \tag{A.111}$$

where the unperturbed Hamiltonian and the perturbation are marked as $\mathcal{H}_0$ and $\mathcal{H}'$, respectively, and the perturbed energies as $\hat{E}_n(\vec{k})$. With this, standard second-order time-independent perturbation theory gives

$$\hat{E}_n(\vec{k}) = E_n(0) + \langle u_{n,0}| \mathcal{H}' |u_{n,0}\rangle + \sum_{n' \neq n} \frac{\langle u_{n,0}| \mathcal{H}' |u_{n',0}\rangle \langle u_{n',0}| \mathcal{H}' |u_{n,0}\rangle}{E_n(0) - E_{n'}(0)}. \tag{A.112}$$

Here, $E_n(0)$ is the unperturbed energy of band n at the Γ-point. Since the expansion is around the point $\vec{k} = 0$, the first order term vanishes due to symmetry.

A situation where $\vec{k} \cdot \vec{p}$ is useful is to examine the behavior around the direct band gap of a solid. Here, two bands, the valence and conduction bands are close to each other and far from other bands, and the energy separation at the Γ-point $E_n(0) - E_{n'}(0)$ is E_g, leading to the full second-order energy expression

$$E_n(\vec{k}) = E_n(0) + \frac{\hbar^2 \vec{k}^2}{2m} + \frac{\hbar^2}{m} \sum_{\alpha,\beta} k_\alpha k_\beta \frac{|\langle v| p_\alpha |c\rangle \langle c| p_\beta |v\rangle|}{E_g}, \tag{A.113}$$

where the valence and conduction band states have been relabeled as v and c. Finally, using the definition Eq. (A.103), the effective mass reads

$$\left(\frac{1}{\mathbf{m}_n^*}\right)_{\alpha,\beta} = \frac{\delta_{\alpha\beta}}{m} + \frac{2}{m^2} \frac{|\langle v| p_\alpha |c\rangle \langle c| p_\beta |v\rangle|}{E_g}, \tag{A.114}$$

which can now be seen to depend directly on the momentum matrix elements. Finally, it is important to note, that due to the inverse proportionality of the effective mass to the optical transition matrix elements, a small mass results in strong coupling, but a small mass also results in a small density of states (DOS), due to the $(\mathbf{m}_n^*)^{3/2}$-proportionality of the DOS, reducing the overall absorption.

A.4.4 Phonon Contribution

If a solid – for simplicity, a crystal – contains more than one type of atom, there is a charge transfer from one type of element to its nearest neighbors of the other sort. This follows from the mere fact that no two elements have the exact same electronegativity. This charge transfer leads to partially to fully ionic bonds within the solid. Vibrations of the atomic (ionic) lattice are called phonons, two types of which exist in solids that are made of a lattice with a basis. One are the acoustic modes where the basis atoms of the same unit cell oscillate in phase, and the other are the optical phonons, where the basis atoms oscillate 180° out of phase. If the basis atoms have a differential charge associated with them, because they are of different kinds, the optical phonon modes correspond to oscillating dipoles and are therefore driven by electromagnetic radiation, they are "infrared active".

This situation is virtually identical to the starting position Eq. (A.55) of the discussion that lead to the Drude-Lorentz Model Eq. (A.60), except now the oscillating particles are not electron shells with rigid nuclei, but two different types of ion cores. The moving mass thus is not the electron mass m but the reduced mass of the two ion cores $M = (1/M_1 + 1/M_2)^{-1}$. The "spring constant" D is now due to the electrostatic interaction between nearest neighbors and the damping constant b, in this simple picture, is due to the anharmonicity of the nearest-neighbor interaction. The resulting interaction once again is a Drude-Lorentz Oscillator,

resulting in a dielectric function of the shape [161]:

$$\varepsilon_r(\omega) = \varepsilon_r^{\text{other processes}} + \frac{e^2}{\varepsilon_0 M v} \frac{1}{(\omega_0^2 - \omega^2) - i\gamma_0\omega}, \tag{A.115}$$

where the electron density N was substituted with $1/v$ the volume of the unit cell. The Drude-Lorentz oscillator in its simple classical form is far more useful for phonons than it is for the electronic oscillations for three reasons. First, due to the greater mass of the ions compared with the electrons, the characteristic frequencies of phonon oscillations $\omega_0 = \sqrt{\frac{D}{M}}$ are 100-1000 times smaller than the ones of electronic oscillations. They are therefore spectrally well isolated from most other contributions. Second, the damping term $\gamma = \frac{b}{M}$ is also generally very small compared with electronic damping terms due to the large reduced mass, and even more so for single crystals due to their generally small lattice anharmonicities. Third, since the ion cores are well defined and localized vibrating masses, and the bonds make rather easily interpretable "springs", the classical picture is generally quite well suited.

This fact makes local field corrections similar to the discussion leading to Eq. (A.65) even more important here. After some calculation, whereby the contributions of the electronic polarizabilities of the different ion cores to the local field are also considered, one arrives at the true resonance frequency ω_T to be substituted for ω_0 into Eq. (A.115):

$$\omega_T = \omega_0 \sqrt{\frac{\varepsilon_r^\infty + 2}{\varepsilon_r^0 + 2}}, \tag{A.116}$$

where ε_r^0 is the static dielectric constant and $\varepsilon_r^\infty = \varepsilon_r^{\text{other processes}}$ is the value of the dielectric function due to other processes, which is valid for $\omega \gg \omega_T$, but still far below any other resonances, i.e. in the mid to near-infrared range. With these two limiting values for the dielectric function, Eq. (A.115) can be rewritten to give

$$\varepsilon_r(\omega) = \varepsilon_r^\infty + \frac{(\varepsilon_r^0 - \varepsilon_r^\infty)\,\omega_T^2}{(\omega_T^2 - \omega^2) - i\gamma\omega}, \tag{A.117}$$

which is just a statement of the fact that for $\omega \gg \omega_T$ the phonon contribution is negligible, because of the inertia of the ion cores, and for $\omega = 0$ the static dielectric constant ε_r^0 is recovered.

It can be shown that the resonance frequency ω_T is the frequency of the transverse-optical phonon modes at the center of the Brillouin zone and the value of ω for which the real part of ε_r vanishes is equal to the frequency of the longitudinal-optical phonon modes at the center of the Brillouin zone. Between these two values, the real part of ε_r is negative, and thus the extinction coefficient k is large and no wave can propagate in the solid. This range is associated with a large reflectivity, and is called the *Reststrahlen* band.

A.4.5 Other Contributions

There are a number of other contributions to the absorption that all fall into one of the following categories. Dopants can give rise to absorptions such as donor to conduction band, valence band to acceptor, acceptor to conduction band, and valence band to donor, and excitonic absorptions. Other crystal defects can give rise to several different types of color centers. In amorphous materials, the interactions can often be described quite analogously to crystals, but, quite phenomenologically speaking, most sharp features are "smeared out" due to the variable environments of the absorbing structures giving rise to a statistical distribution of the absorption energies. The Brendel oscillator model discussed later in this chapter is a good dispersion model for such situations.

In amorphous materials, often there tend to be voids that can absorb different contaminants from the atmosphere. One of the main contributors here is water vapor, which has its absorption lines in the infrared region reducing the transparency there.

Molecular solids, especially organic solids have their molecular vibrations that might not be well described by phonons, but rather by the picture of free molecules. Nearly all organic molecules have bond vibrations, for instance the C-O bond that lie within the infrared, reducing transparency there. That is the main reason, aside from their thermal instability, why this work completely focusses on inorganic materials for infrared optical coatings.

Permanent dipoles in liquid phase materials such as water, can be made to rotate at low frequencies, as happens inside a microwave oven. This drastically changes the optical properties of such materials at low frequencies.

Birefringence occurs when the polarizability of a material is non-isotropic. This leads to a non-isotropic refractive index.

Other contributions include scattering of light at periodic structures, for instance as they appear dynamically when the crystal vibrates, i.e. phonons. This gives rise to inelastic scattering processes, such as Raman scattering that creates or annihilates a phonon during the scattering of a photon. Of course there is also scattering at rough interfaces such as microcrystals within the solid.

Finally, non-linear processes arise when the light field is strong enough to give rise to second-order, i.e. multiphonon effects.

A.4.6 Overview of Dielectric Functions Over a Wide Spectral Range

This section discusses some features of experimental data of a few materials compiled from the literature to give an overview of how the different contributions of the previous chapters interact to give the full spectrum of the dielectric functions. Figure A.4 gives a qualitative overview of how the imaginary part of the dielectric function ε_2 behaves over four decades of the electromagnetic spectrum on the example of a doped III-V semiconductor at low temperatures. Clearly visible is the background of free carriers modeled by the Drude term starting off at approximately 300 cm^{-1} and decaying to 1 cm^{-1} at about 1eV. On top of that there are distinct absorption peaks, to be modeled very well with few Drude-Lorentz

Oscillators, for the donor and acceptor ionizations and their recombinations as well. The transverse optical phonon absorption is shown as a very strong peak around 30-40 meV. At just a little over 1eV, where the photonic energy is not quite enough to lift an electron from the valence band to the conduction band, there is an excitonic peak and just above that, at the band gap, is the onset of the broad interband transitions, modeled as a continuum of overlapping oscillators.

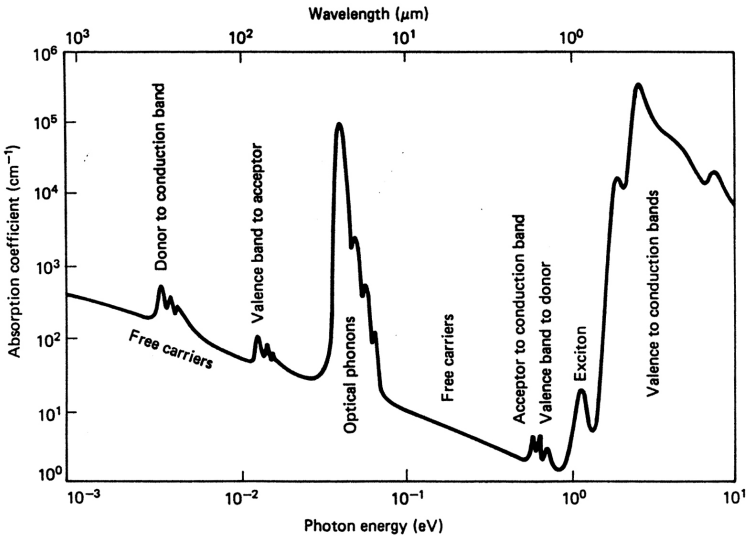


Figure A.4 *Overview of the loss spectrum of a hypothetical III-V semiconductor. Figure from [161]*

Figures A.5 – A.7 show literature data of the real and imaginary parts of the first Group IV solids diamond (red), silicon (green), and germanium (blue) for comparison. These are all purely covalent semiconductors that crystallize in the diamond structure with sp^3-bonds. Figure A.6 shows well how the onset of electronic absorption increases the lighter the atom is, a consequence of the screening effect of the growing number of core electrons. Figure A.5 shows well how the the very different band gap of these materials is mostly responsible for the decrease in the static dielectric constant in the order Ge, Si, D according to Eq. (A.75). This can be seen in Fig. A.5: Diamond has a static $\varepsilon_1 = 5.8$ and an index of refraction of $n = 2.4$, silicon has $\varepsilon_1 = 11.7$ and $n = 3.4$, and germanium has $\varepsilon_1 = 16.0$ and $n = 4.0$. Therefore the index of refraction in the infrared range is solely determined by far-lying interband transitions.

It is interesting to note however, that the onsets visible in Fig. A.6 do not correspond to the band gaps of the materials. This is most noticeable for diamond, whose band gap is at 5.5eV, although the plot suggests approximately 7eV. This is because the effect of the band gap onset is too small and the scale too coarse in Fig. A.6. In the log plot of Fig. A.7 the band gaps of diamond at 5.5eV, silicon at 1.1eV, and germanium at 0.7eV are clearly visible, each with a small, barely visible excitonic bump just before them.

Instead, the visible onset of the absorption of Fig. A.6 followed by extraordinarily large values of ε_2 is a consequence of a great Joint Density of States (JDOS) of two or more bands. The JDOS grows for bands that run parallel to each other in k-space, because the condition for direct transitions is met for large ranges of k-vectors. This is obvious from Fig. A.8 on

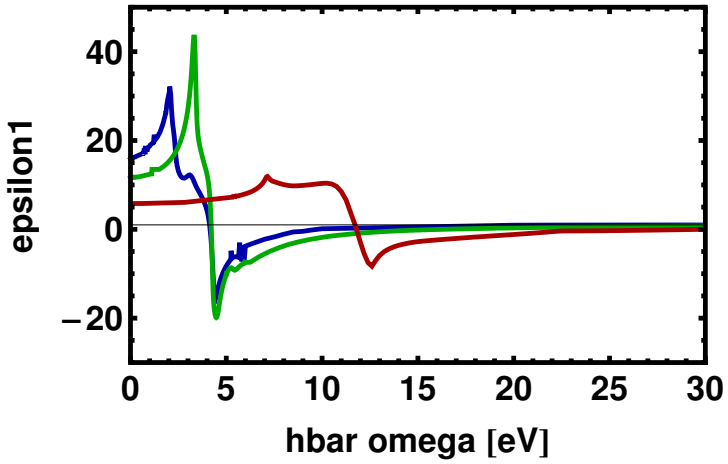


Figure A.5 *Literature data of the real part of the first Group IV solids diamond (red), silicon (green), and germanium (blue) for comparison. Data from [188]*

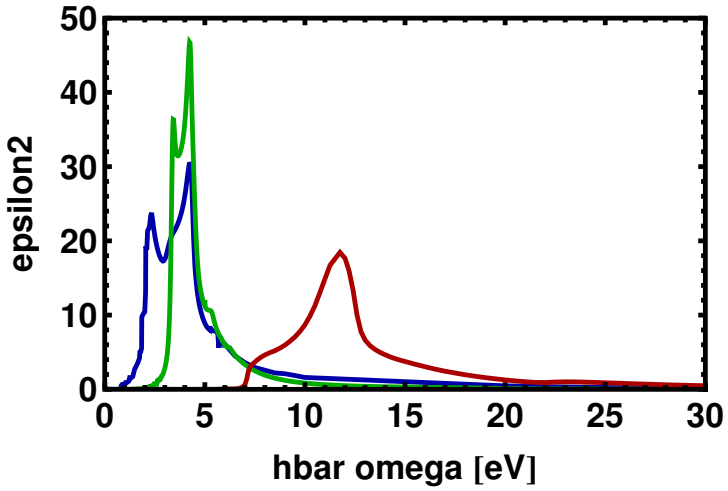


Figure A.6 *Literature data of the imaginary part of the first Group IV solids diamond (red), silicon (green), and germanium (blue) for comparison. Data from [188]*

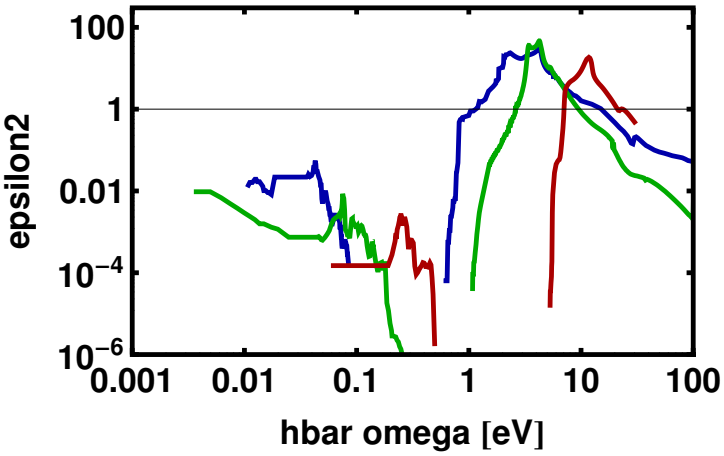


Figure A.7 *Log Plot of the literature data of the imaginary part of the first Group IV solids diamond (red), silicon (green), and germanium (blue) for comparison. Data from [188]*

the example of silicon, where it can be seen, that the indirect band gap of 1.1eV is between the Γ-point and the X-valley, but at a range of values between approximately 3.4 eV, the value of the direct band gap, and 4.2 eV, the valence and conduction bands track each other throughout nearly all of k-space, matched by very large absorption peaks in that range in Fig. A.6.

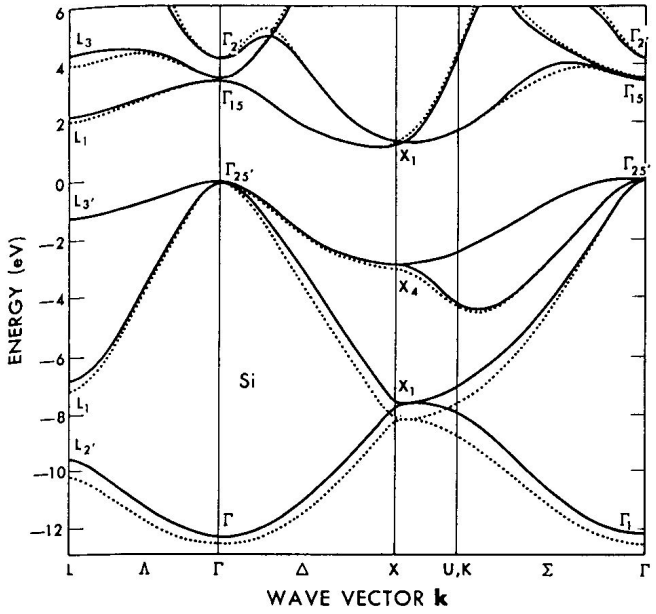


Figure A.8 *Band structure of silicon.*

Figure A.7 also shows the great transparency within the band gap of each material, only disturbed by impurities between 0.2 and 0.5 eV for diamond, 0.05 and 0.2 for silicon, and 0.02 and 0.1 eV for germanium, as well as some intrinsic free carriers below that for silicon and germanium. These make for very small absorption though. Although the diamond structure

of all three materials contains optical phonons, these are not infrared active, since the basis is made of the same type of atoms. They are therefore not shown in these figures. Figure A.5 shows how the oscillations of ε_1 coincide with the peaks of ε_2 in Fig. A.5 and how below the interband transitions, ε_1 falls steadily to its constant static values without undergoing any more oscillations. It also shows how above the interband transitions there is a range where $\varepsilon_1 < 0$, thus giving rise to strong reflection due to a large value of $k \approx \sqrt{|\varepsilon_1|}$. At even higher frequencies, in the deep ultra-violet and X-ray ranges, the material becomes transparent again with $0 < \varepsilon_1 < 1$ (the horizontal line indicates 1). In this range the phase velocity inside the medium is faster than in the vacuum.

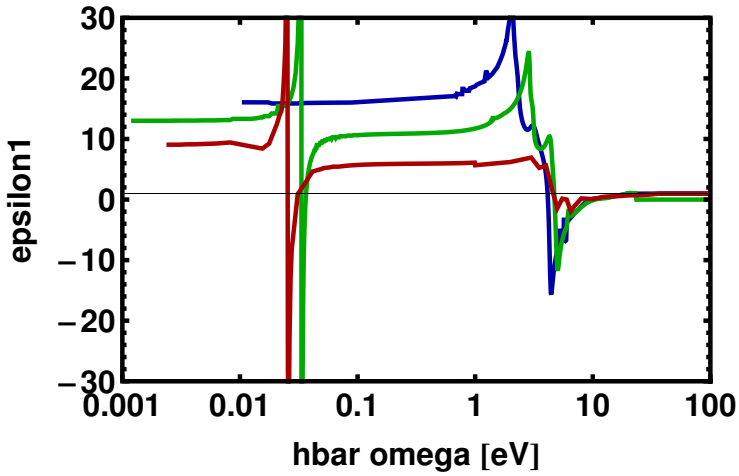


Figure A.9 *Real part of the dielectric functions for the Period IV semiconductors centered around germanium. Blue: Germanium. Green: gallium arsenide. Red: zinc selenide. Data from [188]*

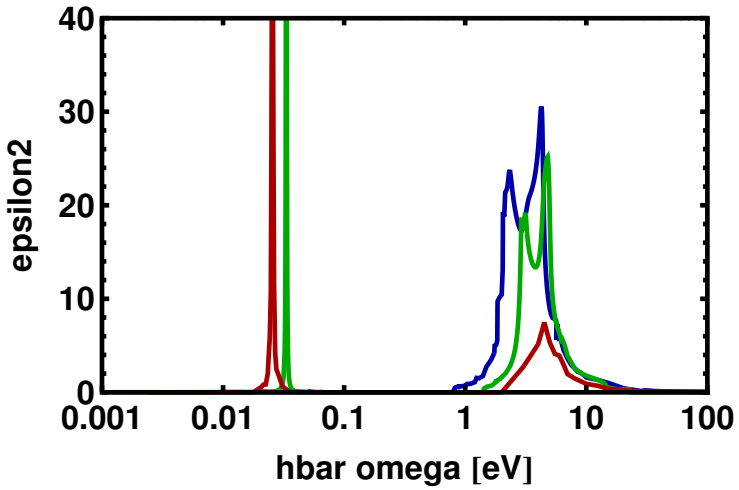


Figure A.10 *Imaginary part of the dielectric functions for the Period IV semiconductors centered around germanium. Blue: Germanium. Green: gallium arsenide. Red: zinc selenide. Data from [188]*

Another interesting group of solids to compare is the Period IV semiconductors centered

around germanium. Germanium is a covalent Group IV semiconductor, gallium arsenide is a III-V compound and zinc selenide a II-VI compound, each of which share the same number of electrons per atomic pair and crystallize in the diamond structure, which for diatomic compounds is called the zincblende structure. The only major chemical difference between these solids is the growing ionicity in the order Ge-GaAs-ZnSe. Figures A.9 and A.10 show the real and imaginary parts of the dielectric functions for these compounds. The first difference to note are the large phonon peaks of GaAs (green) and ZnSe (red) in Fig. A.10, not for Ge (blue) though, because its phonons are not infrared active as stated before. Since phonons are deformations of the lattice, their energy is related to the overall deformability of the solid, thus the harder it is, i.e. the greater Young's modulus, the higher their energy. Ge, GaAs, ZnSe have TO-Phonon energies of 300.0, 268.7, and 207.0 eV, respectively, while they have Young's moduli of 103, 85.5, and 67.2 GPa.

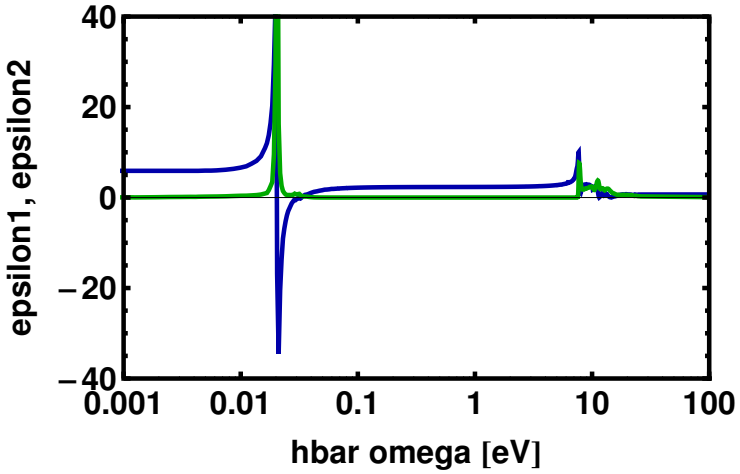


Figure A.11 *Real (blue) and imaginary (green) parts of the dielectric functions of sodium chloride. Data from [188]*

It is also interesting to note, that Fig. A.10 clearly shows a reduction of overall electronic absorption with increasing ionicity. This is even more obvious in Fig. A.11 of an extreme case of an ionic material, sodium chloride, where the electronic absorption is barely eve visible next to the phononic peak. The reason for this is once again the band structure of the material. The more ionic the material is, the better the valence electrons are localized around the cores. In a tight-binding picture, well-localized electrons correspond to flat bands. But if the valence band is flat, but the (empty) conduction band has no physical reason to have a very flat band, then these bands do will not track each other, resulting in a small JDOS. This is the quantum-mechanical reasoning behind the classical notion that small, hard, and spherical ions are not easily polarizable. Thus, due to the Kramers-Kronig relations, the more ionic a material is, the smaller the refractive index at the same band gap. This is only true for the high-energy side of the optical phonons though, since the stronger the phonon absorption, the higher the refractive index on their low-energy side due to the same reason. Therefore, in the mid-infrared region between the phononic absorption and the electronic absorption, the

more localized the electron cloud is, the smaller the refractive index.

Finally, another interesting thing to note about Fig. A.11, is the fact that the sharp peak starting the onset of the interband transitions is due to excitons. But contrary to Fig. A.7, the peak here is well visible and stronger even than the interband transitions. This is due to the fact that the very low ability of a hardly polarizable ionic material to screen electrons and holes from each other results in very small, i.e. well-localized excitons. But well-localized in real space means spread out in k-space. And finally, since now both the valence band and excitonic level are essentially flat, they are also parallel and have a large JDOS resulting in large absorption.

B

Ellipsometry

Ellipsometry is a powerful experimental tool to investigate the optical properties of materials to assess their suitability in thin film optical coatings. It is also a very accurate tool to characterize the structure of the coatings in terms of layer thicknesses, interface roughnesses, layer homogeneities, and so on. It was used extensively throughout this thesis for the characterization of the effects of the process parameters on the optical properties as well as the calibration of the deposition systems in use.

Ellipsometry measures the change of the polarization state of light reflected from the surface of a sample at non-normal incidence. The polarization state changes because components of the electric field vector parallel and perpendicular to the plane of incidence on a surface have different complex Fresnel amplitude reflection coefficients. This change in polarization allows conclusions to be drawn as to the optical properties of the reflecting medium. In the case of one or more layers deposited on a substrate, it is the multiple internal reflections within the layers that yield information about optical properties and thicknesses of all the layers and the substrate, presuming the layers are sufficiently smooth and transparent for the light to penetrate to the substrate. Ellipsometry is a suitable tool to analyze films or stacks of dielectrics, semiconductors, even metals if sufficiently thin, on transparent or absorbing substrates of any sort. Standard texts providing thorough information are available [189, 190].

Spectroscopic ellipsometry (SE), as opposed to single-wavelength ellipsometry (SWE), determines this change of polarization state over a range of wavelengths allowing the optical dispersion relations of the constituting layers to be studied as opposed to their values at a specific wavelength only. SE has some substantial advantages over reflectance measurements with respect to stability and sensitivity [191]. Furthermore, since the change in polarization as a complex number is determined, it is two quantities measured per wavelength as opposed to one (the reflected intensity) in reflectometry, eliminating some ambiguities of reflectance measurements [190](page 229). While reflectometry is usually performed at normal or near-normal angles of incidence, ellipsometry has another degree of freedom since its measurements are customarily performed at multiple angles of incidence allowing for even greater amounts information to be gathered about the sample.

B.1 Setup of an Ellipsometric Measurement

Figure B.1 shows a schematic configuration of an ellipsometric measurement, Fig. B.2 shows a photograph of the SENDIRA Spectroscopic Infrared Ellipsometer by Sentech Instruments GmbH. One arm prepares light to a certain well-defined state and emits it towards a sample, which in turn reflects it towards the second arm, to analyze the resultant state. For the sake of simplicity in this first step, let the source emit monochromatic unpolarized light and ignore the existence of the compensator. Refer to Fig. B.1 for the coordinate systems in use. There are two of them, one before, the other after reflection off the sample. The x-axes are in the plane of incidence, that also contains the sample normal. The y-axes are collinear and perpendicular to the plane of incidence. The following steps calculate the field amplitude seen by the detector in dependence of the relevant parameters of the setup. This derivation follows the line of argument given in [190].

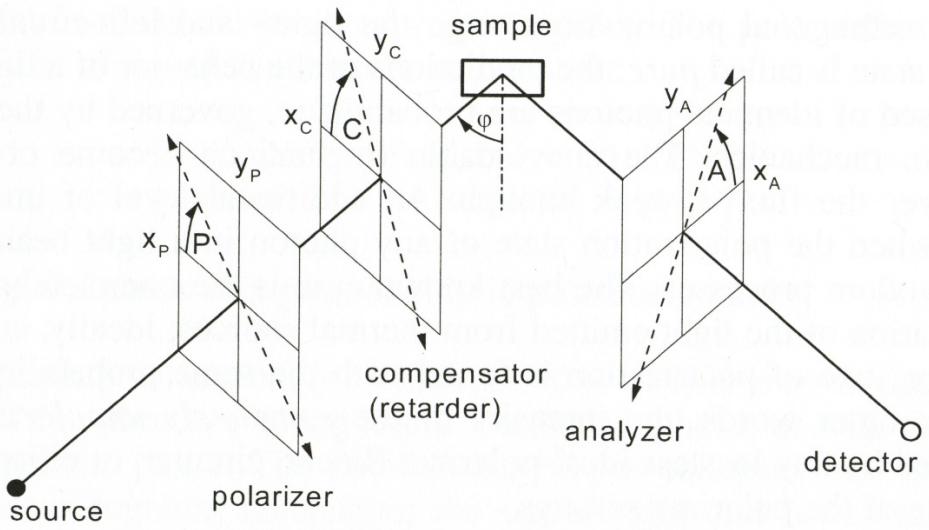


Figure B.1 *Polarizer-compensator-sample-analyzer (PCSA) configuration of an ellipsometer. Image taken from [190].*

1. The polarizer has the effect of introducing a well-defined linear polarization on the unpolarized light from the source. The polarization axis lies in the x-y-plane at an angle P with respect to the x-axis in positive direction of rotation. Let E_P be the complex amplitude of the linearly polarized light transmitted by the polarizer, this amplitude is independent of the angle P if the source in fact emits unpolarized light. (In the following, we will call the polarizer element and its azimuth, i.e. the axis itself, as well as the angle with the x-axis, all by the same name P. The exact reference is clear from the context. The same holds for the compensator and analyzer.)

2. In general, light incident on the sample will have components parallel and perpendicular to the plane of incidence, these are called p- and s-polarizations, where 's' stands for the German *senkrecht*. The reflected wave is a superposition of the p- and s-components

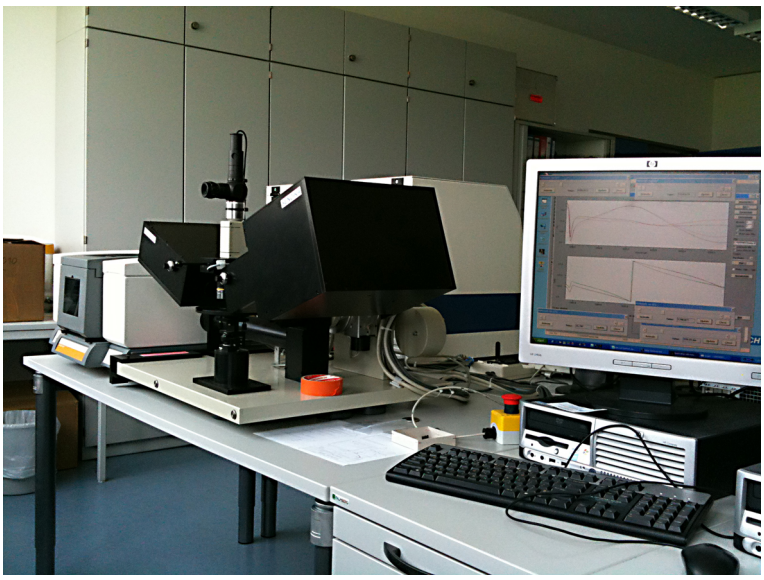


Figure B.2 *SENDIRA Spectroscopic Infrared Ellipsometer by Sentech Instruments GmbH.*

that are modified by multiplying them with their respective reflectivities r_p and r_s. The amplitudes therefore are

$$\begin{aligned} E_{rp} &= r_p E_P \cos P \qquad \text{and} \\ E_{rs} &= r_s E_P \sin P \end{aligned} \tag{B.1}$$

3. The analyzer is nothing but a second polarizer with a polarization axis in the x-y-plane at an angle A rotated in positive direction from the x-axis. The name *analyzer* stems from the fact that in conjunction with the detector it is used to analyze the polarization of the light reflected from the sample. Light transmitted by the analyzer is the sum of the p- and s-polarized components of the light reflected off the sample, each projected onto the polarization axis A. This gives

$$\begin{aligned} E_A &= E_{rp} \cos A + E_{rs} \sin A \\ &= E_P (r_p \cos P \cos A + r_s \sin P \sin A) \\ &= E_P r_s (\rho \cos P \cos A + \sin P \sin A) \end{aligned} \tag{B.2}$$

Where ρ is the ratio of the reflectances r_p and r_s. These are generally complex quantities, therefore ρ is as well. As such it can be expressed in terms of two real quantities, the modulus and phase. For reasons that will be clear later this section, the modulus is better expressed as the tangent of a real angle. Let

$$\rho = \frac{r_p}{r_s} = \tan \Psi e^{i\Delta} \tag{B.3}$$

The angles Ψ and Δ are called the *ellipsometric parameters* and are the central observables in an ellipsometric measurement, their significance will become clear soon. Eq. (B.3) is called the *fundamental ellipsometric equation*. Consider a fixed azimuth P and a variable azimuth A for the following. The physical observable to be measured by the detector is not the amplitude of the electric field E_A but the intensity, which is proportional to $|E_A|^2$. Thus with the substitution of Eq. (B.3) into Eq. (B.2) the intensity is

$$\begin{aligned} I(A) &= I(P)|r_s|^2 \cos^2 P(\tan^2 \Psi \cos^2 A + \tan^2 P \sin^2 A \\ &\quad 2 \tan \Psi \cos \Delta \tan P \cos A \sin A) \end{aligned} \tag{B.4}$$

The analyzed state of the reflected wave is obviously independent of the intensity of the light source. Eq. (B.4) therefore has three independent unknowns, $I(P)$, the intensity after the polarizer P, Ψ and Δ. Thus, the elimination of $I(P)$ and the determination of Ψ and Δ require the measurement of the intensities at three different settings of the azimuth A. Careful choice of these simplifies the math. If measurements are performed at $A = 0$, $A = \pi/4$, and $A = \pi/2$ the ellipsometric angles are given by

$$\tan \Psi = |\tan P| \sqrt{I(0)/I(\pi/2)} \tag{B.5}$$

$$\cos \Delta = \mathrm{sgn} P \cdot \frac{2I(\pi/4) - I(0) - I(\pi/2)}{2\sqrt{I(0)I(\pi/2)}} \tag{B.6}$$

Why is the method termed *ellipsometry*? An answer to this combined with some insight into the physics can be gained by the following considerations. Linearly polarized light hits the sample. Its components parallel and perpendicular to the plane of incidence undergo different reflections determined by their respective reflectivities. The reflections alter these components in amplitude and generally introduce a phase change between them (for simplicity, decoherence effects upon reflection are ignored at the moment). Thus, the resulting light is no longer linearly but elliptically polarized, owing to the fact that the most general case of polarization is elliptic. This ellipsis is now measured in shape and orientation by the analyzer, thus allowing conclusions to be drawn on the reflectivities of the sample, and in consequence, the inherent optical properties of the sample that result in these reflectivities.

More precisely, adding the time-dependence to Eq. (B.1), the reflected wave has the electric field vector (e.g. in the plane at $z = 0$):

$$\vec{E}_r(t) = E_P \begin{pmatrix} r_p \cos P \\ r_s \sin P \end{pmatrix} e^{-i\omega(t-t_0)} = E_P r_s \begin{pmatrix} \rho \cos P \\ \sin P \end{pmatrix} e^{-i\omega(t-t_0)}. \tag{B.7}$$

Assume P to be oriented at 45°. Then

$$\vec{E}_r(t) = E_P \frac{r_s}{\sqrt{2}} \begin{pmatrix} \rho \\ 1 \end{pmatrix} e^{-i\omega(t-t_0)} = E_P \frac{r_s}{\sqrt{2}} \begin{pmatrix} \tan \Psi e^{i\Delta} \\ 1 \end{pmatrix} e^{-i\omega(t-t_0)}. \tag{B.8}$$

And finally,

$$\vec{E}_r(t) = E_P \frac{r_s}{\sqrt{2}\cos\Psi} \begin{pmatrix} \sin\Psi e^{i\Delta} \\ \cos\Psi \end{pmatrix} e^{-i\omega(t-t_0)}. \tag{B.9}$$

The part before the parentheses on the right hand side of Eq. (B.9) determines the overall intensity, the part in the parentheses is the normalized Jones vector of the polarization state. The angles Ψ and Δ have geometric interpretations and result in an ellipse as is seen in Fig. B.3. The endpoint of the electric field vector precesses along the elliptic trajectory, one revolution is achieved in a time of $2\pi/\omega$. Since Ψ and Δ are real numbers, it can be seen from Eq. (B.9), that the y-component is at its maximum for $t = t_0$ and the x-component after a time interval of Δ/ω, denoted by the circumscribing rectangle. Therefore the angle between the dashed and dotted arrows in Fig. B.3 is related to Δ. Special cases of elliptic polarization are linearly polarized light, for which $\Delta = 0$ or π, and circularly polarized light for which $\Psi = \pi/4$ and $\Delta = \pi/2$ or $-\pi/2$.

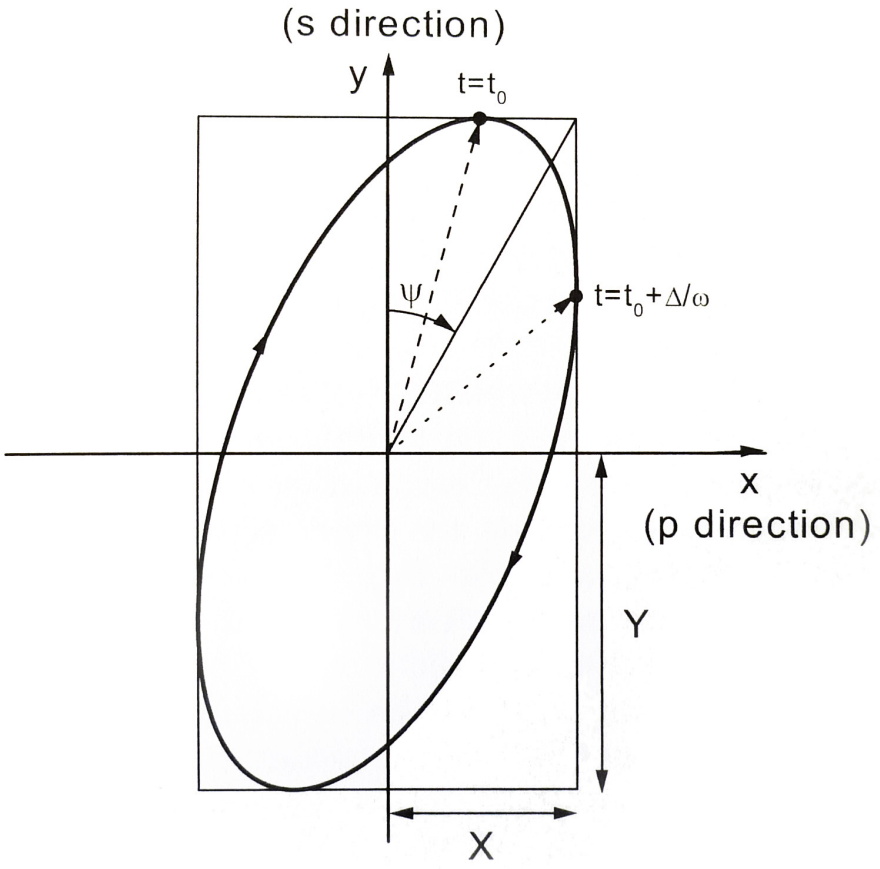


Figure B.3 *Polarization ellipse, described by the ellipsometric angles Ψ and Δ. The polarized wave propagates in the positive sense of the z-axis, which points towards the reader. Image taken from [190].*

The reflectivities r_p and r_s are the Fresnel reflectivities, Eqs. (2.15) and (2.21), for the simple case of a bare substrate only. In the case of any layers present, the reflectivities are to be understood as the stack reflectivities, Eq. (2.17). In this case the ellipsometric equation (B.4) are not simply invertable to give the optical properties of all the layers present. This

problem is dealt with in Appendix B.2.

There is an apparent disadvantage in Eq. (B.6), in that only the cosine of Δ is given, therefore its sign remains unknown. This flaw can be overcome by inserting the compensator shown in Fig. B.1. A compensator or retarder is a birefringent plate with a well-defined thickness, introducing a phase change between the two orthogonal components of the transmitted light. This is, because it has different refractive indices for light polarized parallel or perpendicular to its own optical axis, these are therefore termed the fast and slow axes. Let both axes be in the x-y-plane, the fast axis be at an angle C with respect to x. The equivalent of Eq. (B.1) is:

$$\begin{aligned} E_{rp} &= r_p(E_{fast}\cos C - E_{slow}\sin C) \qquad \text{and} \\ E_{rs} &= r_s(E_{fast}\sin C + E_{slow}\cos C) \end{aligned} \tag{B.10}$$

where

$$E_{fast} = t_{fast}E_P\cos(P-C) \qquad \text{and} \qquad E_{slow} = t_{slow}E_P\sin(P-C) \tag{B.11}$$

from similar geometric arguments that lead to the derivation of (B.2), t_{fast} and t_{slow} are the transmissivities of the retarder plate to components polarized parallel to the fast and slow axis, respectively.

The case discussed in this section is the highly ideal case of perfectly polarized, i.e. *pure* states of light, with no introduced decoherence by either the polarizers or the sample. Neither was there polarization sensitivity or noise in the detector or source and the angular settings of the optical elements contained no errors. The great advantage of ellipsometry in the real world lies in the fact, that through the large number of degrees of freedom for the same measurement, all the non-ideal parameters are sufficiently over-determined to perform minimization of errors through regression. For instance, the ellipsis can be measured at many more than three angular settings of A. The polarizer P, the compensator C, and the angle of incidence ϕ can be varied as desired. More on this can be found in standard texts such as [189–191]. Maybe the greatest potential lies in the connection of all this information measured continuously in a wide range of wavelengths. This method is termed *spectroscopic ellipsometry* and its features are dealt with in section B.2.

Finally, it is illustrative to discuss as an example a simple case of a measurement. Take the dependence of Ψ and Δ of the angle of incidence ϕ. In the following, for clarity, we will rewrite Eq. (B.3) as

$$\rho = \tan\Psi e^{i\Delta} = \frac{r_p}{r_s} = \frac{\tan\Psi_p}{\tan\Psi_s}e^{i(\Delta_p-\Delta_s)} \tag{B.12}$$

Here we have written each of the complex reflectivities as a modulus and phase, although the geometric interpretation as an ellipse given earlier does not apply here, because the phase and amplitude change here compares either component to its value before the reflection, thus the rays are spatially separated.

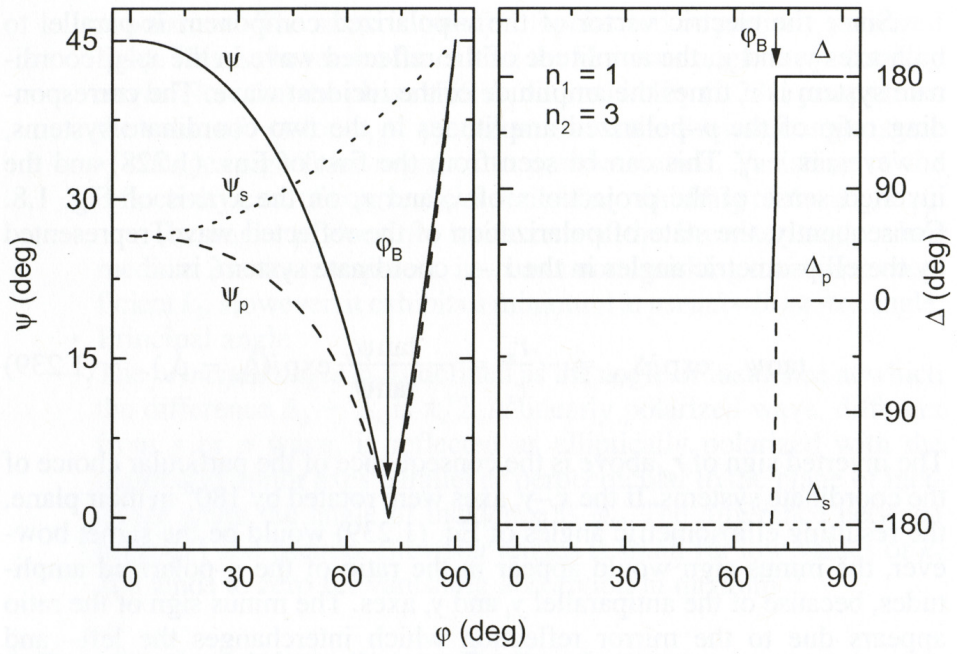


Figure B.4 *The ratios (solid lines) of Fresnel amplitudes for p-polarized (dashed lines) and s-polarized (dotted lines) waves as functions of the angle of incidence for the interface between two non-absorbing media ($n_1 < n_2$). The arrows point to the Brewster angle. Image taken from [190].*

Consider the reflection on an non-absorbing bare smooth substrate in air. Refer to Fig. B.4 for this. At $\phi = 0$ parallel and perpendicular components are physically indistinguishable, thus their values Ψ_p and Ψ_s are equal and $\Psi = 45°$. At the Brewster angle, r_p vanishes and with it Ψ_p and Ψ. Below the Brewster angle, for both components the electric field of the incident and reflected rays are phase-shifted by π as can be seen in the right graph. At the Brewster angle, the phase of the parallel component jumps from $-\pi$ to 0, thus the difference Δ jumps from 0 to π. The effect of non-zero absorption is to smoothen the kink and discontinuity at the (pseudo-)Brewster angle.

B.2 Data analysis

In this section, the data analysis for the general case of a stack of individual layers on top of a substrate is discussed. Though in the present experimental setup there was only one layer deposited on a substrate, the general case is used for the analysis, to allow for deviations of the ideal case in terms of surface and interface roughnesses as well as inhomogeneities in the form of composition gradients over sputter time.

If more than the bare substrate is ellipsometrically measured, i.e. there are deposited layers present, the Ψ and Δ values at a given wavelength cannot be simply inverted to give values for n and k and the thickness of each layer [191]. A regression analysis is therefore required. The analysis starts with setting up a physical model of the layer stack, consisting of a substrate, ambient (air), and the layers in the correct order and some guessed thicknesses. Additionally,

models to calculate the dispersion relations $n_i(\lambda)$ and $k_i(\lambda)$ of each layer material i are included. The models can be either a simple polynomial, e.g. a Cauchy dispersion model, or a model based on approximations to some physical mechanism, e.g. Drude-Lorentz or Brendel Oscillator models. See section B.2.2 for details on this. These models use a certain number of free parameters, which, along with the layer thicknesses, are fitted with a regression algorithm to match the experimental data. For that, all parameters assume guessed seeding values upon which the values for Ψ and Δ of the stack for every given wavelength are calculated and compared to the measured ones with the aid of a merit function, usually the Mean Square Error (MSE). Through a numerical algorithm, e.g. a Levenberg-Marquardt algorithm, the MSE is minimized, yielding the best values for the thicknesses and free parameters of the dispersion models, allowing the dispersion relations of all layers to be calculated. Obviously, the overall number of free parameters cannot exceed the number of measured values. Certain standard materials, like the silicon substrate and air have well-tabulated optical properties that are usually not fitted in order to reduce the number of free parameters and with that the room for errors.

B.2.1 Errors and development of accurate physical models

There are a number of different sources affecting the accuracy of the results and the errors on those. The quantities measured are Ψ and Δ of the layer stack for each wavelength. The quantities of interest though, are the optical properties n and k (over a range of wavelengths) and the thickness of each layer individually.

The experimental errors of Ψ and Δ are a combination of random and systematic instrumentation errors. The random errors are due to detector noise, but due to the nature of an ellipsometric measurement, not the intensity fluctuations of the light source, as opposed to e.g. a reflectometry measurement [191]. The noise is minimized by integrating over a number of individual measurements and using sufficiently smooth functions in the regression analysis to approximate the measured curves. The systematic errors are due to instrument calibration errors, e.g. deviations from the nominal angle of incidence, polarizer angle settings, wavelength de-calibration of the spectrometer, etc. The angle of incidence is also not precisely defined but spread out over a certain range of angles, due to the fact that the instrument focuses the light to a spot on the sample. These systematic errors are challenging to quantify, but are fortunately dominated by the errors of the regression.

The second class of systematic errors stem from the regression analysis. Unfortunately, the MSE between the calculated Ψ and Δ curves and those measured as an indicator for the closeness of the result, i.e. calculated optical constants and thicknesses, to the 'truth' is to be seen with much care. A number of cases have to be considered. For one, parameters can converge to wrong values if the actual measurement is insensitive to these, even if the qualitative model is accurate. An example of this is a thick layer of metal on a substrate. If it is thick enough to attenuate most of the light before it hits the substrate interface, the measurement will be unable to yield correct results for the actual thickness. Two, in

the unlikely case that two different sets of physical models yield the same Ψ and Δ, the MSE cannot distinguish between these even if it converges to a very small value. Generally speaking, a wrong physical model will yield wrong results regardless of the actual value of the MSE. Care has to be taken when devising a model, accounting for interfacial roughnesses, lateral inhomogeneities, composition gradients, etc. Finally, it is not trivial to tell whether the MSE converges into the global or a local minimum within the limits of the parameters, even if the physical model is carefully devised.

The usual procedure comprises the following steps. First, to devise a model as carefully as possible, followed by a regression run. If the minimization of the MSE is successful, i.e. has reached the global minimum, its absolute value and the correlation behavior of the free parameters, given by the correlation matrix, yield the errors of each of them. Too large a correlation between two or more parameters is an indicator for at least one of them being redundant and to be eliminated. Too large an MSE value is a sign of the model being qualitatively inaccurate, with certain aspects or properties being poorly guessed. In the next step, a new model is derived from the old one and the regression re-run. This is to be done iteratively until the overall MSE and the uncertainties of the parameters are satisfactory. These uncertainties subsequently translate into the error bars of the optical property functions and thicknesses.

B.2.2 Physical models (layer stacks) and Dispersion models

The regression algorithm calculates Ψ and Δ for each wavelength from the complex amplitude reflection coefficients $r_{parallel}$ and $r_{senkrecht}$ which in turn are calculated for the layer stack using a matrix method as introduced by Weinstein in 1947 [128] and discussed in Section 2.2. This method solves the boundary value problem of the E and B field within the numerous layer interfaces when an electro-magnetic wave of a given wavelength λ impinges on the stack. Since it accounts for the amplitude reflections and transmissions at the interfaces and the phase shift during translation through the individual layers i, it requires the optical thickness of all layers, i.e. the complex index of refraction $\tilde{n}_i(\lambda) = n_i(\lambda) + k_i(\lambda)$ and the layer thicknesses d_i. The values for $n_i(\lambda)$ and $k_i(\lambda)$ in turn are calculated from model equations that account for material properties through adjustable parameters, although in some simple cases these equations are mere polynomials with no physical significance.

To assess the suitability of a given material for use in optical coatings the optical properties for a variety of wavelengths in the MIR need to be measured. This is done by fitting open (semi-)empirical parameters in model dispersion relations for the layers to the overall Ψ and Δ functions measured for the samples to yield the correct dispersion relations for each material. Since the real and imaginary parts of ϵ as well as $\tilde{n}$ are not independent of each other but linked by the Kramers-Kronig Relation, if one is known over a sufficiently broad region, the other can be directly calculated.

Different mathematical models are used to approximate n and k. The simplest being a constant e.g. for air $n = 1$ and $k = 0$. Others either model n and k empirically like the

Cauchy model, or model the dielectric function in a more or less sophisticated manner from first principles, from which n and k can be calculated. Others take the density of states of the materials into consideration to calculate the imaginary part of the dielectric function, from which the entire dielectric function is then derived by Kramers-Kronig integration.

Appendix A.3 contains a thorough review of the optical properties of solids. One particular dispersion model, termed the *Brendel oscillator model* used to fit the experimental results of the ellipsometric measurements performed throughout this thesis is described in the next section.

The result of our thorough investigation of suitable materials for AR coatings has been published published earlier [192].

B.2.3 The Brendel Oscillator Model

The Brendel Oscillator Model is a versatile tool used to model absorption bands in (partially) amorphous solids and is the primary model used in this work in the MIR and in the visible range. A Brendel oscillator is an inhomogenously broadened Drude-Lorentz oscillator, i.e. the number of Drude-Lorentz oscillators per frequency interval around a center frequency are governed by a Gaussian distribution. The inhomogeneous broadening allows for a further degree of freedom attributing to partial amorphousness of the absorbing material and therefore yields a better fit with less parameters. A dielectric function with Brendel oscillators each of which are a convolution of a Gaussian function with Lorentz oscillators, was designed to fulfill the Kramers-Kronig-Relation as well. Thus, while the Drude-Lorentz model has the form

$$\epsilon_{DL}(\nu) = \epsilon_\infty + \sum_{j=1}^{m} X_j(\nu) \tag{B.13}$$

where

$$X_j(\nu) = \frac{\nu_{pj}^2}{\nu_{0j}^2 - \nu^2 - i\nu_{\tau j}\nu}, \tag{B.14}$$

the Brendel oscillator model has the same form, with the oscillators having the form

$$X_k(\nu) = \frac{1}{\sqrt{2\pi}\sigma_k} \int_{-\infty}^{\infty} \exp - \left(\frac{(x-\nu_{0k})^2}{2\sigma_k^2} \right) \frac{\nu_{pk}^2}{x^2 - \nu^2 - i\nu_{\tau k}\nu} \mathrm{d}x, \tag{B.15}$$

where σ_k is the standard deviation of the Gaussian distribution. The ratio $\nu_{\tau k}/\sigma_k$ determines the shape of the imaginary part from purely Drude-like to purely Gaussian.

As with the Drude-Lorentz model, there are two different schools of thought when assigning Brendel oscillators to absorption features in the measured Ψ-Δ curves. One is to use as few oscillators as can be supported by physical arguments, i.e. with real vibrational transitions, whether or not the Ψ-Δ curves can actually be well approximated by these only. Another approach is to use as many oscillators as needed to fit the spectra well, even though this may lead to some strictly speaking unphysical oscillators. In defense of this approach it can be

argued that the n and k functions of the material, be it pure or contaminated by traces of other substances, will be more accurate. Also unphysical oscillators can be used to attribute to sharp features in the joint density of states that overlay some broad absorption lines, giving the impression that other transitions are present. Since this work is only interested in giving the most accurate n and k values possible, the latter approach is chosen.

C

Numerical Programs

C.1 Photon Density Propagation Model

An implementation of the model discussed in the Section 2.5.4, which runs on Wolfram Mathematica version 8 is stated in the following along with comments in the code. It reads a P-I curve from a .txt file, computes a set of P-I pairs for the same current values from the model with the use of the parameter set and the shooting method. It then compares the simulated curve with the experimental one by computing a figure of merit that is the sum of the relative errors squared. This figure of merit is then minimized by finding the best set of parameters using the Nelder-Mead method and a given number (5) of Random Seed sets. The final output is the simulated curve plotted against the experimental one along with the best set of parameters.

```
(*read P-I curve, where power values have been converted to photon flux
values for ease of use*)

b = ReadList["bare.txt", Number,
   RecordLists -> True];

(*values for phis, xi r2 numerically stabilized :
phis-> phis/10, xi-> xi*10000, r2->r2*100,r1->r1*100*)

(* self-consistent solution: phi1(x), phi2(x) for a given set of parameters
i, xi, g, alpha, phis*)

s1[i_?NumericQ, xi_?NumericQ, g_?NumericQ, alpha_?NumericQ,
   phis_?NumericQ] := NDSolve[{
    phi1'[x] ==
     phi1[x] (g i/(1 + (phi1[x] + phi2[x])/(phis*10)) - alpha) + (xi/
```

```
        10000) i,
    phi2'[x] == -phi2[x] (g i/(1 + (phi1[x] + phi2[x])/(phis*10)) -
        alpha) - (xi/10000) i,
    phi1[0] == Abs[(r1/100) phi2[0]],
    phi2[l] == (r2/100) phi1[l]}, {phi1, phi2}, {x, 0, l},
  Method -> {"Shooting",
    "StartingInitialConditions" -> {phi1[0] == 14.5,
      phi2[0] == 14.5/(r1/100)}}, AccuracyGoal -> 3,
  PrecisionGoal -> 4];

(*output photon flux density to the left*)

f1[i_?NumericQ, xi_?NumericQ, g_?NumericQ, alpha_?NumericQ,
   phis_?NumericQ] :=
  Evaluate[phi1[l] (1 - (r2/100)) /. s1[i, xi, g, alpha, phis ]][[1]];

(*output photon flux density to the right*)

f2[i_?NumericQ, xi_?NumericQ, g_?NumericQ, alpha_?NumericQ,
   phis_?NumericQ] :=
  Evaluate[phi2[0] (1 - (r1/100)) /. s1[i, xi, g, alpha, phis ]][[1]];


(*table that lists the simulated output photon flux to the left for every
current value from the experimental list*)

bsim[xi_?NumericQ, g_?NumericQ, alpha_?NumericQ, phis_?NumericQ] :=
  Table[{b[[i ]][[1]], f1[b[[i ]][[1]], xi, g, alpha, phis]}, {i,
    Length[b]}];

(*fix some of the parameters*)
r1 = 28;
r2 = 28;
l = 0.6;
alpha = 4;


(*the figure of merit is the sum of the relative errors squared
((P_meas-P_sim)/P_meas)^2*)
```

```
figofmerit [phis_?NumericQ, xi_?NumericQ, g_?NumericQ] :=
  Total[((Transpose[b][[2]] -
        Transpose[bsim[xi, g, alpha, phis ]][[2]])/
      Transpose[b][[2]])^2];

(*this minimizes the figure of merit, thus fitting the parameters
phis, g, xi to the experimental P-I curve*)
Do[
 NMinimize[{figofmerit[phis, xi, g], 2 <= phis <= 15, 0.5 <= g <= 3,
   0.5 <= xi <= 10}, {{phis, 4, 8}, {g, 1, 1.8}, {xi, 4, 5}},
  StepMonitor :>
   Print[figofmerit[phis, xi, g], ", ", phis, ", ", g, ", ", xi],
  AccuracyGoal -> 3, PrecisionGoal -> 2,
  Method -> {"NelderMead", "RandomSeed" -> i}], {i, 5}]


(*with the determined values for phis, xi, and g the simulated and experimental
P-I curves can be plotted together*)

phisres = 12.563457745418551` ;
xires = 50.5041021379274895`;
gres = 1.3739071999704604`;

LP = ListPlot[{b}];
LPsim = ListPlot[{bsim[xires, gres, alpha, phisres]}, Joined -> True];

Show[LP, LPsim]
```

C.2 Design and Simulation of Interference Filters

```
(* initializing  variables*)
Clear[lambda, a1, a2, a3, d1, d2]


(*this reads a measured transmission spectrum to plot along the
simulated curve for comparison*)

measured = ReadList["transmission.txt", Number, RecordLists -> True];

(*build the plot*)
LP = ListPlot[{measured}, PlotRange -> {0, 0.6}];

(*material values for the index of refraction at the central
wavelength*)
ns = 3.42; (*correct value*)
ny2o3 = 1.686 - I 0.0023;
nsi = 3.42 - I 0.00006;(*+I 0.02;*)
nal2o3 = 0.73218 - I 0.83855;
nsin = 1.8334;
nsio2 = 1.376;
nsio = 1.78;
nair = 1;
nge = 4.025;
nau = 2.941 - I 26.45;
ny2o3x = 1.8458;
nznse = 2.432;
nmgf2 = 1.35 - I 10^(-7);
ncaf2 = 1.40 - I 10^(-7);
nyf3 = 1.488 - I 0.00045;


(*----------insert parameters here----------*)
\
(*----------------------------------------*)

designwavelength = 4.0(*4.394*); (*in microns*)
```

```
angle = 0*\[Pi]/
    180; (*10 Grad in Rad*)(*angle of incidence*)
designangle =
 45*\[Pi]/180;(*angle of incidence for central wavelength*)

Np = 3;(*number of pairs*)

nL = nyf3;(*low index*)
nH = nge;(*high index*)
nSpacer = nyf3;
nInit = ny2o3;(*initial AR coating*)

(*thicknesses of high and low-index layers*)
dH =
 designwavelength/(4 Re[nH]) /Cos[ArcSin[Sin[designangle]/Re[nH]]] -
  0.0024
dL = designwavelength/(4 Re[nL]) /
    Cos[ArcSin[Sin[designangle]/Re[nL]]] - 0.0024
(*thickness of spacer*)

dSpacer =
  0 designwavelength/(2 Re[nSpacer]) /
    Cos[ArcSin[Sin[designangle]/Re[nSpacer]]];
(*thickness of antireflection coating*)

dInit = designwavelength/(4 Re[nInit]) /
    Cos[ArcSin[Sin[designangle]/Re[nInit]]];

(*thickness of substrate*)
ds = 400;(*in microns*)

(*ordering of the material indexes in pairs*)
nPair1 = {nH, nL};(*starting next to substrate, stack1 closer to substrate*)

nPair2 = {nL, nH};(*starting next to substrate, stack2 closer to ambient*)

(*build the DBRs*)
nStack1 =
 Table[nPair1[[Mod[i, 2] + 1]], {i, 0,
    Np*2 - 1}];(*starting next to substrate*)
```

```
nStack2 =
 Table[nPair2[[Mod[i, 2] + 1]], {i, 0,
   Np*2 - 1}];(*starting next to substrate*)

(*build the stack*)

nLayers = Join[{nInit, nsi, nInit}, nStack1, {nSpacer}, nStack2];


(*ordering of the layer thicknesses in pairs*)
dPair1 = {dH, dL};
dPair2 = {dL, dH};

(*build the DBRs*)
dStack1 =
 Table[dPair1[[Mod[i, 2] + 1]], {i, 0,
   Np*2 - 1}];(*starting next to substrate*)
dStack2 =
 Table[dPair2[[Mod[i, 2] + 1]], {i, 0,
   Np*2 - 1}];(*starting next to substrate*)

(*build the stack*)

dLayers = Join[{0.7 dInit, ds, dInit}, dStack1, {dSpacer}, dStack2];


(*build the stack*)
nAll = nLayers;
PrependTo[nAll, nair];
AppendTo[nAll, nair];
Do[n[i] = nAll[[i + 1]], {i, 0, Length[nLayers] + 1}];

(*assign snell's angles*)

Do[phi[i] = ArcSin[nair/Re[n[i]] Sin[angle]], {i, 0, Length[nLayers]}];
phi[Length[nLayers] + 1] = angle; (*einfallswinkel*)


phiLayers = Table[phi[i], {i, Length[nLayers]}];
phiAll = Table[phi[i], {i, 0, Length[nLayers] + 1}];
```

```
deltaLayers = 2 \[Pi]/lambda nLayers dLayers Cos[phiLayers];
Do[delta[i] = deltaLayers[[i]], {i, Length[deltaLayers]}];


(* parallel component*)
etaPAll = nAll/Cos[phiAll];
Do[etaP[i] = etaPAll[[i + 1]], {i, 0, Length[etaPAll] - 1}];
MP[i_] := {{Cos[delta[i]],
    I/etaP[i] Sin[delta[i]]}, {I etaP[i] Sin[delta[i]],
    Cos[delta[i]]}};
x0P = {1, etaP[0]};
xP = x0P;
Do[xP = MP[i].xP, {i, Length[nLayers]}];
EmP = xP[[1]];
HmP = xP[[2]];
rP = (etaP[Length[nLayers] + 1] EmP -
     HmP)/(etaP[Length[nLayers] + 1] EmP + HmP);
tP = (2 etaP[Length[nLayers] + 1])/(etaP[Length[nLayers] + 1] EmP +
     HmP);
RP = Abs[rP]^2;
TP = etaP[0]/etaP[Length[nLayers] + 1] Abs[tP]^2;



(*senkrecht component*)
etaSAll = nAll Cos[phiAll];
Do[etaS[i] = etaSAll[[i + 1]], {i, 0, Length[etaSAll] - 1}];
MS[i_] := {{Cos[delta[i]],
    I/etaS[i] Sin[delta[i]]}, {I etaS[i] Sin[delta[i]],
    Cos[delta[i]]}};
x0S = {1, etaS[0]};
xS = x0S;
Do[xS = MS[i].xS, {i, Length[nLayers]}];
EmS = xS[[1]];
HmS = xS[[2]];
rS = (etaS[Length[nLayers] + 1] EmS -
     HmS)/(etaS[Length[nLayers] + 1] EmS + HmS);
tS = (2 etaS[Length[nLayers] + 1])/(etaS[Length[nLayers] + 1] EmS +
     HmS);
```

```
RS = Abs[rS]^2;
TS = etaS[0]/etaS[Length[nLayers] + 1] Abs[tS]^2;


lambda = 10000/wn;
range = designwavelength; (*tunability range*)

Print["Reflection, Transmission Coefficient Rp, Tp[lambda]"]
Plot[{Rp , Tp}, {wn, 1, 2000}, PlotRange -> {0, 1}]

Print["Reflection Coefficient Rs[lambda]"]
Plot[{RS}, {wn, 500, 6000}, PlotRange -> {0, 1}]

Print["Reflection, Transmission Coefficient Rp, Tp[lambda]"]
Plot[{Rp , Tp}, {wn, 880, 920}, PlotRange -> {0, 1}]

Print["Transmission Coefficient Ts[lambda] "]
Plot[{TS}, {wn, 600, 4500}, PlotRange -> {0, 1}]

Print["Transmission Coefficient Ts[lambda] "]
TPlot = Plot[{TS}, {wn, 2260, 2290}, PlotRange -> {0, 1}];

Show[LP, TPlot]
txt = Table[{wn, TS}, {wn, 2260, 2300, 0.25}];

(*write to file *)

Export["normalincidence-fit.txt", txt,
  "Table"];

(* fit central peak*)

fitdata = Table[{wn, N[TS, 3]}, {wn, 2270, 2310, 0.1}];
FindFit[fitdata,
 T (1/2 FWHM)^2/((wn - wn0)^2 + (1/2 FWHM)^2), {{T, 0.8}, {wn0,
   10000/designwavelength}, {FWHM, 2}}, wn]

Print["Layer Thicknesses"]
dLayers
```

```
Print["Overall Thicknesses"]
Sum[dLayers[[i]], {i, Length[nLayers]}]

Print["Phase Values"]
lambda = designwavelength;
deltaLayers
Clear [lambda]
```

C.3 Optimization Program for Anti-Reflection Coatings

```
(* initializing variables *)
Clear[lambda, a1, a2, a3, d1, d2, d3]

(*set values for optical properties at central wavelength*)

ns = 3.25;
ny2o3 = 1.469 + I 0.024;(**)
nsi = 3.42;
nal2o3 = 0.73218 + I 0.83855;
nbi2o3 = 2.2;
nsin = 1.8334;
nsio2 = 1.823 + I 0.049;
nsio = 1.78;
nair = 1;

(*materials array for simpler selection *)

material = {ns, ny2o3, nsi, nal2o3, nsin, nsio2, nsio, nbi2o3, nair};

(*----------insert figures to optimize to here----------*)

designwavelength = 11.2; (*in microns*)

(*define the layer structure*)
layermaterial = {material [[2]],
  material [[8]],
  material [[9]]}; (*starting next to substrate*)
```

```
layerthickness = {d1,
  d2, d3};

(*
weighting parameters for migure of merit:
for the curve R:
minimize Area under curve (a1) + value at designwavelength (a2) + \
symmetry around designwavelength (a3)
        because we want little absorption and little reflection but we dont \
want to use NMaximize for T
*)
a1 = 1;
a2 = 0.5;
a3 = 1;
range = 0.4; (*tunability range*)

(*program start*)

delta = 2 \[Pi]/lambda layermaterial layerthickness;

t[i_] := 2 If[ i <= Length[layermaterial], layermaterial[[i]],
    nair]/(If[ i <= Length[layermaterial], layermaterial[[i]], nair] +
      If[i >= 2, layermaterial[[i - 1]], ns])

r[i_] := (If[ i <= Length[layermaterial], layermaterial[[i]], nair] -
    If[i >= 2, layermaterial[[i - 1]], ns])/(If[
     i <= Length[layermaterial], layermaterial[[i]], nair] +
    If[i >= 2, layermaterial[[i - 1]], ns])

M[i_] := 1/
   t[i] {{E^(-I delta[[i]]),
    r[i + 1] E^(I delta[[i]])}, {r[i + 1] E^(-I delta[[i]]),
    E^(I delta[[i]])}}

x0 = {1, r[1]};

x = x0/t[Length[layermaterial] + 1]; Do[
 x = M[i].x, {i, Length[layermaterial]}];

R = Abs[x[[2]]/x[[1]]]^2; (*Reflection*)
```

```
T =
 ns/nair   Abs[1/x[[1]]]^2; (*Transmission*)
(*A=1-R-T;Absorption*)

layerthickness  = layerthickness  /.  Last[NMinimize[{Hold[

        NIntegrate[
         a1 R + a3 Abs[R (lambda - designwavelength)], {lambda,
          designwavelength - range, designwavelength + range}] +
        a2 With[{lambda = designwavelength}, R]

       ],  {0 <= d1, d1 <= 2, 0 <= d2, d2 <= 2, 0 <= d3, d3 <= 3}},
     layerthickness ]];

(*--------output----------*)
Print[" Sensitivity of R to thicknesses"]
Plot3D[

 With[{lambda = designwavelength, d3 = layerthickness[[3]]}, R],

 {d1,  layerthickness [[1]]  - 0.5,  layerthickness [[1]]  + 0.5}, {d2,
  layerthickness [[2]]  - 0.1,  layerthickness [[2]]  + 0.1},
 PlotRange -> {0, 0.01},
 ColorFunction -> Function[{x, y, z}, Hue[.65 (1 - z)]],
 AxesLabel -> Automatic]


Print["Reflection Coefficient R[lambda]"]
Plot[With[{d1 = layerthickness[[1]], d2 = layerthickness [[2]],
    d3 = layerthickness [[3]]},   R], {lambda, designwavelength - range,
  designwavelength + range}, PlotRange -> {0, 0.002},
 AxesLabel -> {"lambda(microns)", "R"}]

Print["Transmission Coefficient T[lambda]"]
Plot[With[{d1 = layerthickness[[1]], d2 = layerthickness [[2]],
    d3 = layerthickness [[3]]},   T], {lambda, designwavelength - range,
  designwavelength + range}, PlotRange -> {0.90, 1},
 AxesLabel -> {"lambda(microns)", "T"}]

Print["Layer Thicknesses"]
```

```
layerthickness
Print["Phase Values"]
phi = 2 \[Pi]/designwavelength layermaterial layerthickness
```

D

Assembly and Alignment of the V-ECQCL

D.1 Assembly and Alignment

The alignment consists of the following steps:

1. Define optical axis For this, the baseplate B1 is slid into place along the guides of the optical bench A and secured. Only B2-B7 are mounted on B1, the QCL B4 is turned off. The laser A1 is turned on and aimed at the alidade B2, with light passing through both holes. The beam expander A2 is mounted between the Laser A1 and the alidade B2, and adjusted so that light remains passing through both holes of the alidade B2 and the expanded beam has a circular, homogeneously illuminated cross section. The optical axis is now defined with respect to B. The alidade is now removed and the variable grey filter A4 is mounted between A1 and A2. The circular aperture A3 is mounted so that the expanded beam illuminates it fully and positioned so that a homogenous light circle hits the parabolic mirror B5 concentrically.

2. Adjust collimating mirror B5 Set the grey filter to strongest possible intensity reduction. Turn on the light A7 and aim it at the QCL B4. Position the microscope A6 in line with the QCL and focus it on the intra-cavity facet (facing away from the microscope, only visible as an embossment) of the contacted QCL stripe, refer to Fig. D.1 for a view through the microscope. Use 10× magnification. Turn screws 1-5 on B6 until, through the microscope, you see laser light from A1 pass the QCL facet on the side. Use the procedure described in Sec. D.1.1 to align the mirror B5 so that laser light from A1 is focused in the y-z-plane that contains the intra-cavity facet of the QCL, and expands into homogeneous circles if the microscope focus plane is varied in the x-direction. Continue using the procedure to translate this focused spot in the y-z-plane to coincide with the intra-cavity QCL facet (the embossment in the x-y-plane). The mirror is now aligned and B6 can be locked.

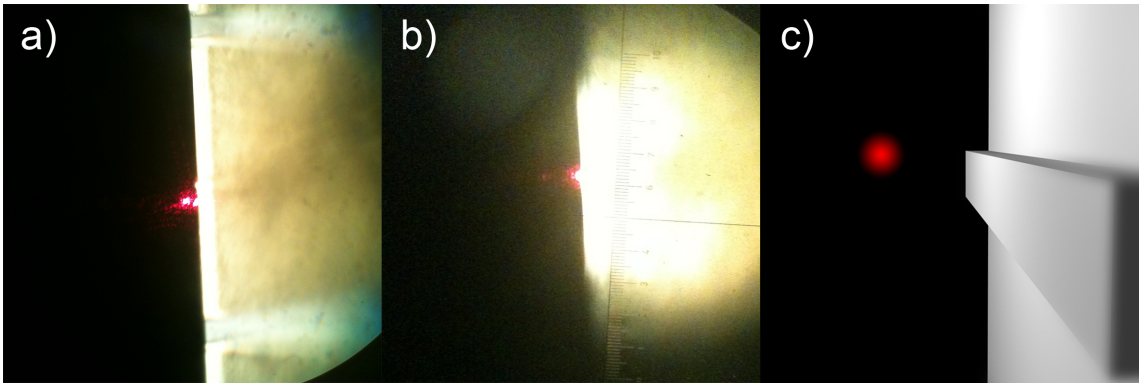


Figure D.1 *The focussed spot (red) as seen through the microscope. a) Spot is correctly focussed on the plane containing the laser facet, visible as a small white embossent, but not yet focussed on the facet. b) Correctly focussed spot on the QCL facet. c) Illustration of a) as a computer-rendered image to clarify what is seen. All three images show a grazing view of the laser ridge on the right. The focal plane contains the laser facet facing* away *from the microscope.*

3. Adjust Grating B9 Remount alidade B2. Assemble grating B9 on mount B8. Mount on rotation stage B7 with the perpendicular of the grating in line with the 0° marking. Rotate grating towards alidade to coarsely adjust grating tilt θ to the optical axis z. Use the top edges of the alidade as reference for the reflection of the laser light A1 from the grating. Now rotate grating to correct first order reflection angle w.r.t. the z-axis facing the mirror B5 according to grating equation $\theta = \arcsin(\lambda/(2d))$ with θ the Littrow angle, λ the center wavelength of the QCL during self-oscillation and d the rule spacing of the grating. This can be a coarse adjustment because this angle can be anywhere within the tuning range of the finished EC-QCL, typical tolerances are $> 3°$. Set up power meter A8 in line with QCL, replacing microscope A6. Turn on power meter and power up QCL to just above lasing threshold. Tune grating tilt δ until power significantly rises. Adjust so that power is maximum. Find repetition rate and pulse current that maximizes output from the QCL with grating feedback, while self-oscillation (with blocked light path towards grating) is small. Now dismount alidade. The External Cavity is set up and can now be detached from the bench A and the laser A1 can be turned off.

4. Collimate output beam (optional) Mount lens B10 on mount B11 and collimate using screws 1-5 on B11 with the aid of Pyrocam A9 or power meter A8 to desired degree.

D.1.1 Mirror Alignment Procedure

Refer to Fig. D.2 a) for nomenclature of actuators. Peer through microscope and observe shape of light from laser A1 ("spot") at 100× magnification. Throughout the entire procedure, turn x-axis knob on microscope forward-backward to witness change of spot shape as the focal plane moves along the light cone, the center position being the plane that contains the QCL facet. The goal of the procedure is 1) to get the center axis of the parabolic mirror (z') parallel to the axis of the collimated laser light (z, "optical axis"), and 2) to move the focal point in the x, y, and z directions to coincide with the laser facet. Due to coma effects, the mirror does not focus all the light into a single spot, if z and z' are not collinear, the spot will be distorted into a straight line or, further misadjusted, an arc. 1) is achieved when the spot is circular and as the focal plane of the microscope moves through the focal point of the parabolic mirror, contracts to a single sharp point, 2) is achieved when this sharp point is directly covered by the laser facet. For an overview of the effects the spot shape as seen through the microscope upon turning each actuator, refer to Fig. D.2 b).

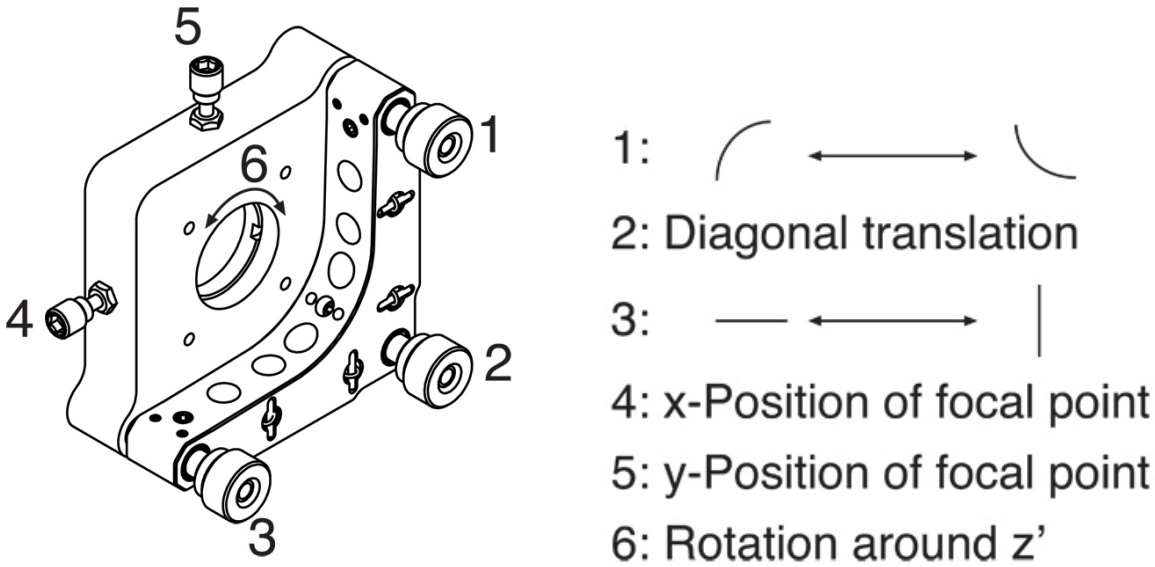


Figure D.2 *Left: 6-Axis actuator with annotated nomenclature for the actuators. Right: Effects on the spot shape as seen through the microscope upon turning each actuator*

1. Turn all actuators to center position coarsely, so that the spot is visible beside the QCL chip
2. Turn mirror around z'-axis until spot is visible beside the facet horizontally (z-direction).
3. Turn actuator 1 to orient elongated spot or arc to be as close to vertical as possible
4. Turn actuator 3 to reduce elongation as the microscope focal plane moves through focus, this is where the line morphs from horizontal to vertical

5. Turn actuator 2 to keep the spot in the field of view of the microscope during steps 3 and 4
6. Repeat steps 3-5 iteratively until the spot is circular and contracts to a single spot at some x position
7. Turn actuator 4 to set the x-position in plane with the QCL facet
8. Turn actuator 5 to set the y-position so that the spot is visible horizontally beside the QCL facet
9. Turn actuators 1,2,3 by the same distance to draw the spot along the x line onto the facet.
10. If at any point the spot becomes distorted, correct the spot shape with steps 3-5.

E

Example of S-ECQCL Software: Performing a quickscan

When the Quickscan button is pressed, an event is triggered and the producer loop fills the "QUICKSCAN" state into the queue to be executed next. The "QUICKSCAN" state is a distributor state, sending the state machine to any of the states "QUICKSCAN 00" to "QUICKSCAN 04" depending on the value of the variable "Scan Stage" that can have an integer value between 0 and 4 and is initialized with the value. It also sets the "Ready" variable connected to the corresponding front panel indicator to FALSE.

Next, in "QUICKSCAN 00", the progress bar is made visible and initialized to the value 0, the "Scan Stage" variable is set to 1 and two states are filled into the queue, the "CLEAR DATA" state, which clears the scan graph on the front panel and the corresponding data caches to make space for the coming scan, and the "QUICKSCAN 01" state.

Next, the "QUICKSCAN 01", state fills the state "GO TO POSITION" into the queue bundled with the number value of the lower scan limit as the targeted position. It also sets the "Scan Stage" variable to 2.

Next, in the "GO TO POSITION" state, the motor movement is initiated, realtime readout of the motors' status bits is initiated, and the state "LOCK FRONTPANEL" is enqueued.

The state "LOCK FRONTPANEL" disallows user input during the Quickscan. This is to avoid changing any of the scan parameters during scan, since this would result in useless spectra. This state fills no further state into the queue, which means the state machine reverts to the "WAITING" state, which is always in the queue and reproduces itself.

The state machine remains in the "WAITING" state, until the "GO COMPLETE" event is triggered by the motor upon arrival at the lower scan limit. This sends the state machine into the state "GO COMPLETE".

In the "GO COMPLETE" state, first the current position of the motor in units of mm is read out and converted to wave numbers using the calibration chart. Next the spectral position is updated in the front panel, along with the progress bar and the step size. The step size in wavenumbers is not a constant, since the actual steps are a multiple of the physical step size of the motor. Next, it is checked, the caller of the movement is checked, whether

it was the program initialization, a scan, a quickscan, or an external trigger. Since it is a quickscan, the “QUICKSCAN” state is filled into enqueued next. The “QUICKSCAN” state immediately enqueues the “QUICKSCAN 02” state because of the “Scan Stage” variable’s value.

Next, in the “QUICKSCAN 02” state, the “Cancel Scan” button is activated and made visible. Next, it is checked, whether the upper scan limit has been reached. If it has not, the queue of the data acquisition SubVI is filled with the “SETUP” state, along with the quickscan parameters to set up the ADC for data acquisition. These parameters differ from the values set on the front panel, since the step size is calculated in such a way that scan time is two minutes and there is a medium integration time for each acquisition. Next, the “GET CONTINUOUS” state is added to the queue of the data acquisition SubVI, which tells the DAC to measure until its cache is filled, send the data back, and immediately start filling the cache again, and do this until further notice. The “GO TO POSITION” state is added to the main queue along with the value of the current position plus one step. The SubVi sends the data back to the main program, by bundling it with the state “DISPLAY” and adding it to the main queue. The “DISPLAY” state then updates the signal monitor and scan graph on the front panel.

This now loops through the motor steps and data acquisition, until the upper scan limit is reached. When the state “QUICKSCAN 02” finds this condition fulfilled, it adds the “STOP GETTING” state into the data acquisition queue, which stops acquisition. It further resets the ADC’s data collection parameters to the values on the front panel (to prepare it for an actual scan and clear the traces of the quickscan process). Finally, it adds the “GO TO POSITION” state with the lower scan limit to the queue and changes the “Scan State” variable to 3.

Upon the next “GO COMPLETE” in conjunction with “Scan State” equalling 3, the state “QUICKSCAN 03” is reached. This clears the quickscan process by hiding the progress bar, unlocking the front panel, and resetting the “Scan State” variable to 0.

F

Official Back Matter

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 14. 04. 2015 Jan-Ferenc Kischkat